ON GEOMETRIES IN WHICH CIRCLES ARE THE SHORTEST LINES*

BY

CARL EBEN STROMQUIST

HAMEL† has considered those geometries in which the straight line is the shortest distance between two points. The axioms at the basis of these geometries are, except for some minor changes, the well known axioms of HILBERT, and are such that by means of them a one to one correspondence can be established between the points in the plane and pairs of numbers (x, y). "Length" of any rectifiable curve $y = y(x)$ between two given points (x_1, y_1), (x_2, y_2) is defined‡ as an integral,

$$\int_{x_1}^{x_2} g(x, y, p)dx, \qquad p = dy/dx.$$

The nature of the geometry will then depend on the particular form of the integral chosen.

HAMEL then studies the geometries for which these axioms are satisfied and for which the integral shall have such a form that its value taken from x_1 to x_2 along a straight line shall be less than that along any other rectifiable curve joining the two points (x_1, y_1) and (x_2, y_2).

HAMEL's methods apply, with but slight modifications, to the problem we wish to consider in this paper; viz., the study of a geometry in the upper half-plane whose elements satisfy the axioms given by HAMEL, but in which the last stated restriction on length shall be replaced by the following:

* Presented to the Society February 28, 1903, under a different title. Received for publication October 28, 1905.

† HAMEL, *Ueber die Geometrieen in denen die Geraden die Kürzesten sind*, dissertation, Göttingen, 1901; in revised form, Mathematische Annalen, vol. 57 (1903), pp. 231–264.

‡ HAMEL defines the length of a straight line between two points as a function of the coördinates of the points and shows that it can be expressed as an integral taken along the straight line. In his dissertation he then tacitly assumes that length along any curve is expressed by the same integral (p. 13); in the article, in the Annalen already referred to. this assumption is definitely stated (p. 241). Another method would be to define length along any curve as the limit of the length of inscribed polygons, as is usual in Euclidean geometry, and then to prove that the limit would be the value of the same definite integral which expressed the length along the straight line segment, taken along the limiting curve. It seems simpler to start with the above definite integral as the definition of length of any curve.

The shortest distance between two points shall be the segment of that circle joining the points whose center lies on the boundary line (x-axis) of the half plane.

This leads to an inverse problem in the calculus of variations. For by this last restriction the shortest lines are given, and the problem is to find the integral

$$\int_{x_1}^{x_2} g(x, y, p)\,dx$$

which will be minimized by these shortest lines. There will in general be an infinite number of solutions to such a problem and our object is to determine the most general function g satisfying these assumptions.

In addition to the problem stated above we shall consider (in section 2) the geometries in which transversals are perpendicular to their extremals; in particular, HAMEL's geometries and the geometries in which circles are the shortest distances. In section 3 examples of particular cases of the "circle geometry" will be given.

§ 1. *Determination of the integrand g.*

To determine the definition of length in the geometries defined above we must then find the most general function $g(x, y, p)$ such that the solution of the problem of rendering

$$l = \int_{x_1}^{x_2} g(x, y, p)\,dx \tag{1}$$

a minimum is given by the function y defined by

$$(x - a)^2 + y^2 = c^2, \tag{2}$$

i. e., the so-called extremals (minimizing curves of our integral) must be the circles with centres on the x-axis, whose differential equation is

$$y'' = \phi(x, y, p) = -\frac{1 + p^2}{y}. \tag{3}$$

From the calculus of variations we know* that the first necessary condition that the curve $y = y(x)$ shall minimize the integral

$$\int_{x_1}^{x_2} g(x, y, p)\,dx$$

is that $y = y(x)$ should satisfy EULER's equation

$$y''\frac{\partial^2 g}{\partial p^2} + p\frac{\partial^2 g}{\partial y\,\partial p} + \frac{\partial^2 g}{\partial x\,\partial p} - \frac{\partial g}{\partial y} = 0.$$

* BOLZA, *Lectures on the Calculus of Variations*, p. 22.

DARBOUX * has pointed out that, conversely, if EULER's equation has as extremals the solutions of $y'' = \phi(x, y, p)$, the function $g(x, y, p)$ must satisfy the differential equation

$$\phi \frac{\partial^2 g}{\partial p^2} + p \frac{\partial^2 g}{\partial y \partial p} + \frac{\partial^2 g}{\partial x \partial p} - \frac{\partial g}{\partial y} = 0. \tag{4}$$

If we differentiate this equation with respect to p we obtain the following linear differential equation for $M = \partial^2 g / \partial p^2$:

$$\frac{\partial M}{\partial x} + p \frac{\partial M}{\partial y} + \phi \frac{\partial M}{\partial p} + M \frac{\partial \phi}{\partial p} = 0, \tag{5}$$

where ϕ is defined by equation (3).

The most general solution of equation (5) in our case is †

$$M = y^2 \cdot W[y^2(1 + p^2), x + yp], \tag{6'}$$

where W is an arbitrary function of the arguments $y^2(1 + p^2)$ and $x + yp$.

It follows that g must have the form

$$g = y^2 \int_c^p \int_c^p W[y^2(1 + p^2), x + yp]\, dp\, dp + p \cdot w_1(x, y) + w_2(x, y), \tag{6}$$

where w_1 and w_2 are functions of x and y alone which must be so restricted that g satisfies equation (4).

Since the differential equation from which g is obtained is equation (4) differentiated with respect to p, the result of substituting g in (4) must result at most in a function of x and y alone. If we call $K(x, y)$ the function resulting from substituting

$$g_0 = y^2 \int_c^p \int_c^p W dp\, dp$$

in (4), then it is readily seen that the result of substituting g in (4) gives

$$K(x, y) + \frac{\partial w_1}{\partial x} - \frac{\partial w_2}{\partial y} = 0.$$

Moreover since the substitution of g_0 in (4) must result in a function free from y' we may evaluate it by giving y' a particular value, say $y' = 0$, from which it fol ows that $K(x, y)$ must have the form $-y \cdot W(y^2, x)$.

Hence our last equation becomes

$$-y W(y^2, x) + \frac{\partial w_1}{\partial x} - \frac{\partial w_2}{\partial y} = 0. \tag{7}$$

* DARBOUX, *Leçons*, vol. 3, p. 53. In solving EULER's equation for our problem we have followed closely the method given by DARBOUX.

† See, for instance, FRICKE, *Analytisch-Funktionentheoretische Vorlesungen*, p. 467.

If $\bar{w}_1$ and w_2 are particular values of w_1 and w_2 respectively, which satisfy (7), then the most general values of w_1 and w_2 are given by

$$w_1 = \bar{w}_1 + \frac{\partial u}{\partial y}, \qquad w_2 = w_2 + \frac{\partial u}{\partial x},$$

where u is an arbitrary function of x and y. We may then choose one of the functions w_1 and w_2 arbitrarily and the other can be determined by equation (7). Let, then, $\bar{w}_2 = 0$, and we have

$$\bar{w}_1 = y \int_{x_0}^{x} W(y^2, x)\, dx.$$

The function g now becomes

$$g = y^2 \int_c^p \int_c^p W[y^2(1+p^2), x + yp]\, dp dp + yp \int_{x_0}^{x} W(y^2, x)\, dx + p \frac{\partial u(x, y)}{\partial y} + \frac{\partial u(x, y)}{\partial x}. \tag{8}$$

This is the most general solution of the problem, where W and u are arbitrary functions of the arguments indicated.

We then know:

1) That the circles (2) are extremals, i. e., they satisfy EULER's equation when g is given by equation (8).

2) That a family of these circles through a point in the upper half plane does not have an envelope in this region, and therefore the condition known in the calculus of variations as JACOBI's condition is satisfied.

If we now impose on g the condition 3) that $g_{pp}(x, y, p) > 0$ for x and y in the upper half plane and for any finite p, the above three conditions are sufficient in order that the arcs of our circles shall render the integral $\int g dx$ a strong minimum as compared to its value for curves in a restricted neighborhood.*

Furthermore, through any two points of the upper half plane we can pass one and only one extremal† [i. e., arc of the circles (2)] and hence this arc will actually render the integral a strong minimum not only in comparison to curves in a restricted neighborhood but in comparison to any curves in our region

* BOLZA, loc. cit., p. 96, theor. V. Also see BOLZA for the definitions of *strong minimum*, *absolute minimum*, *neighborhood*, etc.

† For two points on the vertical we can consider this vertical as the limiting case of our circles when the center has receded to infinity along the x-axis. In this case p is not finite, but this can be overcome by taking as the definition of length in this case

$$\int_{y_1}^{y_2} g\left(x, y, \frac{1}{q}\right) q dy$$

where $q = dx/dy$. Or by taking the problem in parameter-representation the above difficulty would be avoided.

joining the two points, i. e., the circles (2) render the integral an *absolute* minimum.

In order that condition 3) shall be satisfied we must impose the following condition on the function W:

$W[y^2(1+p^2), x+yp] > 0$ for x and y in the upper half plane and for every finite p.

The function g can be reduced to a form containing only one integration, if we make the following transformation:

Let us define the function U such that

$$W[y^2(1+p^2), x+yp] = \frac{\cos^2\theta}{y^3} U(y^2\sec^2\theta, x+y\tan\theta),$$

where $p=\tan\theta$, $c=\tan\theta_0$. Then by integration by parts g can readily be reduced to the form

$$g = \frac{1}{y\cos\theta}\int_{\theta_0}^{\theta} \sin(\theta-\tau)\cdot U(y^2\sec^2\tau, x+y\tan\tau)\,d\tau + \frac{\tan\theta}{y^2}\int_{x_0}^{x} U(y^2, x)\,dx + p\frac{\partial u}{\partial y} + \frac{\partial u}{\partial x}. \tag{9}$$

We have therefore the following result:

Any geometry in the upper half plane in which the element of length is defined by

$$dl = dx\left\{\frac{1}{y\cos\theta}\int_{\theta_0}^{\theta} \sin(\theta-\tau)\cdot U(y^2\sec^2\tau, x+y\tan\tau)\,d\tau + \frac{\tan\theta}{y^2}\int_{x_0}^{x} U(y^2, x)\,dx + p\frac{\partial u}{\partial y} + \frac{\partial u}{\partial x}\right\},$$

where $U>0$ for all x and y in the upper half plane and for every finite p, has the circles $(x-a)^2+y^2=c^2$ as shortest distances.

Any two points* may be joined by one such circle which has a shorter length, according to the definition of length given above, than any other curve joining the two points.

If we impose HAMEL's strong monodrome axiom—that the length of a curve from A to B shall be the same as the length from B to A—we must further restrict g. Following HAMEL† without giving details we obtain the result: In order that the strong monodrome axiom shall be satisfied it is necessary and sufficient that U shall have the period π considered as a function of θ; the values of $\partial u/\partial x$ and $\partial u/\partial y$ are given by

* See last footnote above.

† HAMEL, loc. cit., p. 18.

$$\frac{\partial u}{\partial x} = \frac{1}{2y}\int_{\theta_0}^{\theta_0+\pi} \sin\tau \cdot U(y^2\sec^2\tau,\ x + y\tan\tau)\,d\tau,$$

$$\frac{\partial u}{\partial y} = -\frac{1}{2y}\int_{\theta_0}^{\theta_0+\pi} \cos\tau \cdot U(y^2\sec^2\tau,\ x + y\tan\tau)\,d\tau - \frac{1}{y^2}\int_{x_0}^{x} U(y^2,\ x)\,dx,$$

and are thus determined aside from a constant. The function g then reduces to

$$g = \frac{1}{2y}\int_{\theta-\pi}^{\theta} \sin(\theta-\tau)\cdot U(y^2\sec^2\tau,\ x + y\tan\tau)\,d\tau. \tag{9'}$$

§ 2. *Transversals perpendicular to their extremals.*

Suppose we have a one parameter set of extremals simply covering a portion of the plane. Let p be the function of x and y which defines the direction of an extremal at the point (x, y). Then a transversal,* $y = Y(x)$, is determined by the equation

$$g(x, y, p) + \left(\frac{dY}{dx} - p\right) g_p(x, y, p) = 0, \tag{10}$$

where dY/dx refers to the transversals.

Geometrically stated,† two transversals to the same set of extremals are two curves which intercept on the extremals arcs along which the integral under consideration has a constant value.

If for any set of extremals the transversals are always perpendicular to the extremals then equation (10) must be an identity in x, y, p when for dY/dx is substituted $-1/p$.

From this it follows readily that g must have the form ‡

$$g(x, y, p) = \psi(x, y)\sqrt{1 + p^2}, \tag{11}$$

where ψ is an arbitrary function of x and y alone.

Consider now any geometry in which the extremals are perpendicular to the transversals. The line element must have the form

$$dl = \psi(x, y)\sqrt{1 + p^2}.$$

But by the theory of surfaces § we can always find a surface whose line element has the form

$$ds = \psi(u, v)\sqrt{du^2 + dv^2}.$$

This surface is represented conformally on the plane by the trivial transformation $u = x$, $v = y$. Also, as is evident from the theorem of GAUSS that

* BOLZA, loc. cit., p. 106.

† BOLZA, loc. cit., p. 172.

‡ This was first pointed out by E. R. HEDRICK in a course of lectures on the calculus of variations (1901–1902).

§ DARBOUX, *Leçons*, livre VII, chap. IV.

geodesics are perpendicular to the curves of constant geodesic distances, any geometry obtained by a conformal transformation of a surface on the plane is one in which extremals are perpendicular to the transversals. Therefore,

The necessary and sufficient condition that a geometry be such that extremals are perpendicular to their transversals is that the geometry be obtained by a conformal transformation of some surface upon the plane.

Let us now consider the result of making this restriction that extremals shall be perpendicular to their transversals in the geometries in which the circles (2) are the shortest distances.

We must then have, from (11) and (8),

$$W[y^2(1+p^2), x+yp] = \frac{\psi(x,y)}{y^2(1+p^2)^{\frac{1}{2}}}.$$

In order that this functional relation shall hold we must have *

$$\begin{vmatrix} \frac{\partial}{\partial x}\left(\frac{\psi(x,y)}{y^2(1+p^2)^{\frac{1}{2}}}\right) & \frac{\partial}{\partial y}\left(\frac{\psi(x,y)}{y^2(1+p^2)^{\frac{1}{2}}}\right) & \frac{\partial}{\partial p}\left(\frac{\psi(x,y)}{y^2(1+p^2)^{\frac{1}{2}}}\right) \\ 0 & 2y(1+p^2) & 2y^2p \\ 1 & p & y \end{vmatrix} \equiv 0,$$

or

$$y\psi_x + (y\psi_y + \psi)\,p \equiv 0.$$

Hence

$$y\psi_x = 0 \qquad \text{and} \qquad y\psi_y + \psi = 0.$$

From the second of these equations we see that ψ must have the form $\psi = c/y$, where c is an arbitrary constant, and therefore from (11),

$$g = \frac{c}{y}\sqrt{1+p^2}.$$

This is the geometry obtained by a conformal representation of the pseudosphere on the plane. In fact, the line element on the pseudosphere † is given by $ds^2 = c^2(du^2 + e^{2u}dv^2)$ (where $-1/c^2$ is the constant curvature). By the conformal transformation $v = x$, $e^{-u} = y$ the line element becomes

$$dl^2 = c^2\frac{dx^2+dy^2}{y^2},$$

or the form given above. Therefore,

The most general geometry in which the circles (2) *are the shortest distances and extremals are perpendicular to their transversals is obtained by a conformal transformation of the pseudosphere upon the plane.*

* PASCAL, *Repertorium der Höheren Mathematik*, vol. 1, p. 52.

† DARBOUX, loc. cit., vol. 3, p. 394.

In the case of HAMEL's geometries * g is of the form

$$g = \int_c^p \int_c^p W(p, y - xp)\, dp dp + p \frac{\partial u}{\partial y} + \frac{\partial u}{\partial x},$$

and if the extremals are to be perpendicular to their transversals, we find, proceeding as above, that

$$g = c\sqrt{1 + p^2}, \qquad \text{or} \qquad l = c \int_{x_1}^{x_2} \sqrt{1 + p^2}\, dx.$$

The case $c = 0$ is excluded by our axioms, and for $c = 1$ the geometry is the ordinary Euclidean geometry; $c \neq 1$ gives simply a magnification of the plane, that is, a Euclidean geometry with a different unit of length. We have therefore the following conclusion:

The Euclidean geometry is the most general geometry possible in which the shortest lines are the straight lines and the transversals are perpendicular to the shortest lines.

§ 3. *Examples.*

Let us now consider some particular examples of the geometries in which the circles are the shortest distances.

Example 1. Let $U = \text{const.} = 1$. From equation (9′) we then obtain

$$l = \int_{x_1}^{x_2} \frac{\sqrt{1 + p^2}}{y}\, dx, \tag{12}$$

and the transversals $y = Y(x)$ must satisfy

$$1 + p \frac{dY}{dx} = 0. \tag{13}$$

From the equation of the circles (2) we have $p = (a - x)/y$, and if we seek the transversals corresponding to the extremals through a given point, as $(1, 1)$, we shall readily have

$$p = \frac{x^2 - 2x - y^2 + 2}{2y(1 - x)}.$$

Then equation (13) becomes, after integration,

$$(x - 1)^2 + (y - k)^2 = k^2 - 1, \tag{14}$$

where k is a constant of integration. Hence the transversals are circles with centers, not at $(1, 1)$, but at $(1, k)$.

Since the transversals are circles, in order to construct a transversal of "radius" l_1 for the extremals through $(1, 1)$ we need only determine the values of y above and below $(1, 1)$ on $x = 1$ such that the distance from $(1, 1)$ to

* HAMEL, loc. cit., p. 14.

LIBRARY
OF THE
UNIVERSITY of ILLINOIS

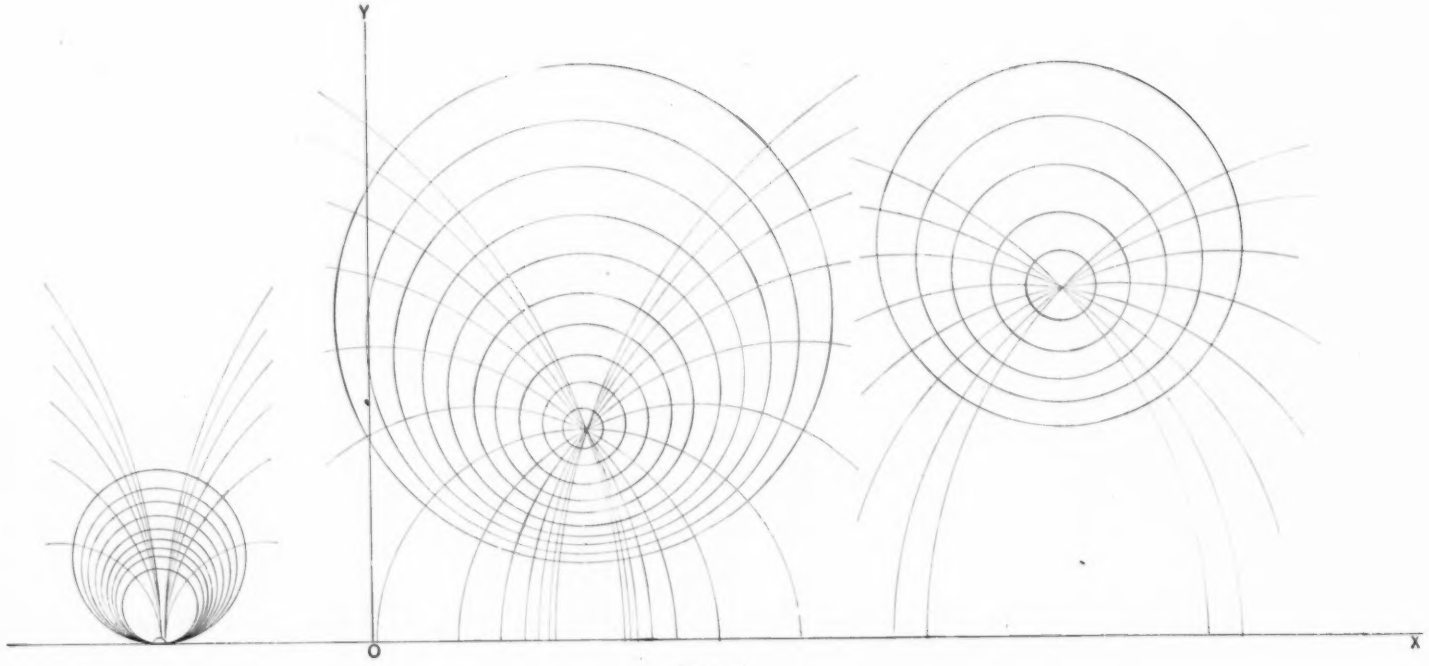


PLATE I.

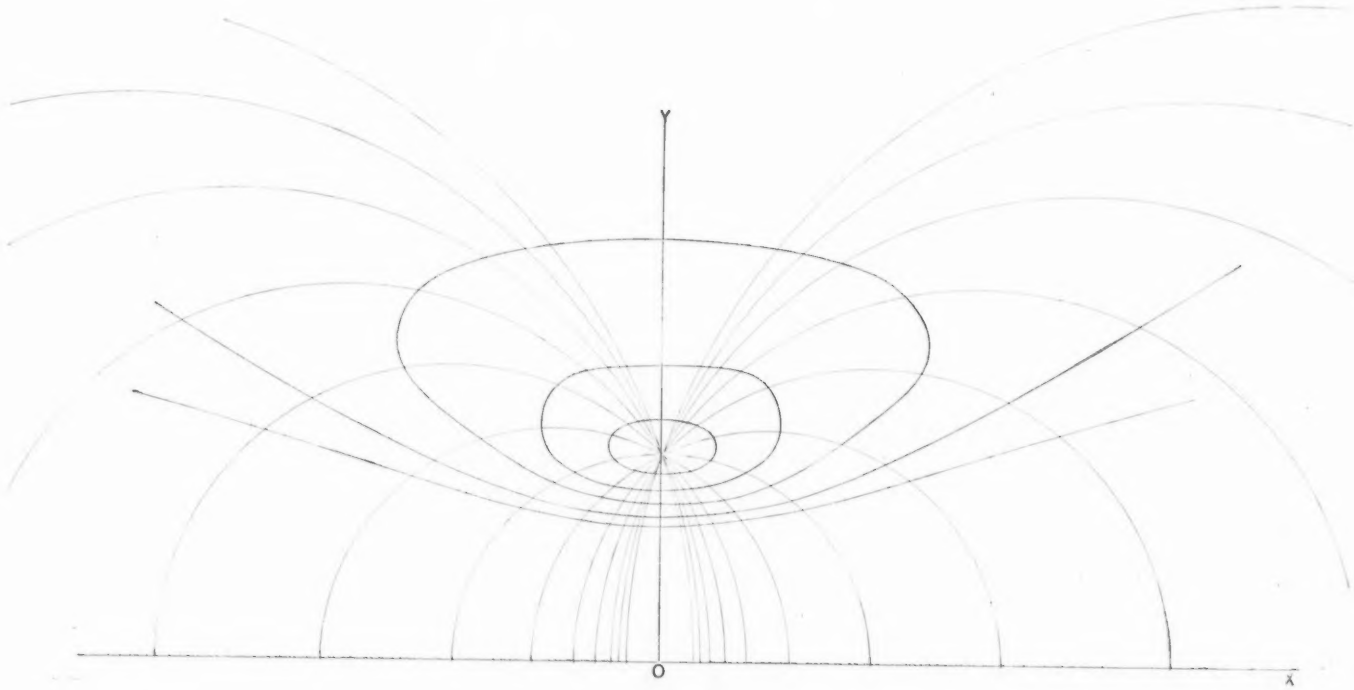


PLATE II.

each of these points is l_1 and then construct the circles (14) through these points and with centers on $x = 1$.

For any line $x = \text{const.}$, equation (12) reduces to

$$l = \log y \Big|_{y_1}^{y_2}$$

which for $y_1 = 1$ and $y_2 = y$ gives

$$l \Big|_1^y = \log_e y. \tag{15}$$

Hence to lay off a given distance l_1 from $(1, 1)$ along $x = 1$ we need only solve for y from $\log_e y = l_1$.

By this method Plate I has been constructed.

Example 2. Let

$$U = \frac{1}{y^2(1 + p^2)} = \frac{\cos^2\theta}{y^2}.$$

In this case we obtain

$$l = \int_{x_1}^{x_2} \frac{1 + 2p^2}{3y^3(1 + p^2)^{\frac{1}{2}}}\, dx. \tag{16}$$

Substituting the values of y and p from the equation of the circles (2) and integrating we have

$$l = \frac{1}{3c}\left[\frac{x - a}{c^2 - (x - a)^2}\right]_{x_1}^{x_2}. \tag{17}$$

Let us construct the transversals corresponding to the extremals through $(0, 1)$. The differential equation for the transversals in this case is somewhat complicated. Instead of attempting to solve it, and to construct the transversals as in the last example, we proceed as follows. If we wish to lay off a length l_1 along any of the extremals (except $x = 1$) from $(0, 1)$, we let $x_1 = 0$, $x_2 = x$ in (17) and solve for x. This gives

$$x = a - \frac{c^2 - a^2}{2\{3cl_1(c_2 - a_2) - a\}} \pm \sqrt{c^2 + \frac{(c^2 - a^2)^2}{4\{3cl_1(c^2 - a^2) - a\}^2}}. \tag{18}$$

By means of this equation we can compute x for any given circle (2) through $(0, 1)$ and for a given l_1.

For $x = \text{const.}$, $dx = 0$ and (16) becomes

$$l \Big|_{y_2}^{y_1} = -\frac{1}{3y^2}\Big|_{y_1}^{y_2},$$

or

$$l \Big|_y^\infty = \frac{1}{3y^2} \qquad (\text{for } y > 0), \tag{19}$$

by which a given length l_1 can be laid off along $x = 1$.

Plate II has been constructed by use of (18) and (19).

PRINCETON, N. J.,
October, 1905.

A GENERALIZATION OF THE NOTION OF ANGLE*

BY

GILBERT AMES BLISS

The tendency of recent writers has been to regard geometry as the science of a certain special kind of assemblage.† The elements of this assemblage, in deference to our intuitive notions called points, fall into subsets called lines. Among the lines and the points of any one line, such properties of intersection, congruence, order and continuity are presupposed as the particular geometer in question is pleased to prescribe, provided only that all the assumptions are mutually compatible. If it is desired that any two geometries satisfying the initial assumptions shall be essentially the same, the axioms must also form a so-called categorical system.‡

The assemblage which forms the basis on which the usual Euclidean and non-Euclidean geometries are built up, is such that its elements can be put into one to one correspondence with the number ratios $x:y:z$, in which (x, y, z) may have any real values different from $(0, 0, 0)$. Starting with assumptions which characterize this assemblage, HAMEL‡ has constructed still more general geometries all sharing the property that straight lines are the shortest distances, and more recently STROMQUIST§ has discussed one in which circles are the shortest.

HAMEL, whose geometries are the most general in which straight lines are shortest, considers the points in his assemblage represented in the usual way on

* Presented to the Society at the Williamstown summer meeting, September 7, 1905. Received for publication October 10, 1905.

† HILBERT considers two related assemblages, one of points and the other of lines; see his *Foundations of Geometry*, translated by E. J. TOWNSEND.

‡ I. e., a system such that the elements of any two assemblages satisfying it can be put into one to one correspondence in such a way that corresponding elements have similar properties; see VEBLEN, Transactions of the American Mathematical Society, vol. 5 (1904), p. 346. The arbitrariness in the choice of assumptions is well illustrated by the geometries in which there are only a finite number of points and lines; see the abstract of a paper by VEBLEN and BUSSEY, Bulletin of the American Mathematical Society, vol. 11 (1905), p. 480.

‡ *Über die Geometrieen in denen die Geraden die Kürzesten sind*, Dissertation, Göttingen, 1901; Mathematische Annalen, vol. 57 (1903), p. 231.

§ These Transactions, vol. 7 (1906), p. 175. An abstract of his paper may be found in the Bulletin of the American Mathematical Society, vol. 9 (1903), p. 396.

an xy-plane and the line at infinity. He then defines the length of a curve to be the value of an integral of the form

$$I = \int f(x, y, y')\, dx,$$

a generalization from the usual definition of length by the integral

$$I = \int \sqrt{1 + y'^2}\, dx.$$

But he omits to show the corresponding generalization of the notion of angle. It is the purpose of this paper to fill in this gap, and to show that the generalization of angle found agrees with the current definitions in the special cases of the non-Euclidean geometries and of geometry on a curved surface. The integral which defines the generalized angle for HAMEL's geometries turns out to be a specialization of an integral invariant connected with a more general problem of the calculus of variations.

In deriving the results of the paper the curves are considered in parametric representation, a form more satisfactory than the other from the geometrical standpoint, and a new form of the definite integral representing length is used which in itself is of interest to the student of the calculus of variations. Its introduction is especially desirable for the reason that the use of the ordinary Weierstrassian form of integral for curves in parametric representation presupposes HAMEL's weak monodromy axiom,* as will be explained in § 1.

§ 1. *The length integral and its extremals.*

Suppose with HAMEL that the points of the geometry to be studied are the xy-points of a plane and the line at infinity, that is, the elements of the geometry are to be in one to one correspondence with the number ratios $x : y : z$, and the ratios with $z \neq 0$ are to be represented in the Cartesian plane. The others constitute the line at infinity.† A curve is then the totality of points defined in the usual way by two equations

$$x = \phi(t), \qquad y = \psi(t), \tag{1}$$

and the length of an arc from $A\,(t = t_0)$ to $B\,(t = t_1 > t_0)$ is defined to be the value of a definite integral

$$l = \int_{t_0}^{t_1} f(x, y, \tau) \sqrt{x'^2 + y'^2}\, dt. \tag{2}$$

In this expression τ is the angle between the x-axis and the tangent to the curve and is defined by the equations

$$\cos \tau = \frac{x'}{\sqrt{x'^2 + y'^2}}, \qquad \sin \tau = \frac{y'}{\sqrt{x'^2 + y'^2}}. \tag{3}$$

*Mathematische Annalen, vol. 57 (1903), p. 244.

† The methods given here apply only to the finite part of the plane. For points on the line at infinity or for curves having such points, a special discussion is necessary.

The integral which WEIERSTRASS introduced into the calculus of variations* is

$$I = \int F(x, y, x', y')\, dt,$$

where the function F is single valued for x', y' not both zero, and positively homogeneous of order one in x' and y'. On account of the homogeneity it can be written in the form

$$I = \int F(x, y, \cos\tau, \sin\tau)\sqrt{x'^2 + y'^2}\, dt,$$

which if used as a definition of length would mean that the function $f(x, y, \tau)$ in (2) would have the period 2π. No such assumption need, however, be made on the function f, and it is desirable not to do so in order to avoid specializing the nature of the geometries which may be derived by HAMEL's method.

In order to carry through the methods of the following pages without going into too much detail with regard to initial hypotheses, it will be assumed that the functions ϕ and ψ have continuous first and second derivatives, and that $f(x, y, \tau)$ has continuous partial derivatives of the first three orders for all values of τ and for values of (x, y) in the neighborhood of the point $O(x_0, y_0)$ at which the generalization of angle is to be made. Furthermore $f(x_0, y_0, \tau)$ and $f + f_{\tau\tau}$† for the same arguments are assumed to be greater than zero whatever τ may be. The last conditions will also be true for (x, y) in a properly chosen neighborhood of the point O provided that τ is restricted to some finite interval.

The differential equation which defines the curves of shortest length or *extremals* for the integral (2), is readily derivable by the method well known in the calculus of variations‡. Suppose that the equations (1) represent such an extremal joining $A(t = t_0)$ with $B(t = t_1)$, and let ξ, η be two functions having continuous derivatives and vanishing at t_0 and t_1. Then the equations

$$C_\alpha: \qquad \bar{x} = \phi(t) + \alpha\xi(t), \qquad \bar{y} = \psi(t) + \alpha\eta(t),$$

represent a set of curves each passing through A and B and reducing to the curve (1) when $\alpha = 0$. The generalized length of the arc AB of C_α has a value depending only upon α, and is expressible in the form

$$l(\alpha) = \int_{t_0}^{t_1} f(\bar{x}, y, \bar{\tau})\sqrt{\bar{x}'^2 + \bar{y}'^2}\, dt,$$

where $\bar{x}, \bar{y}, \bar{\tau}$ have the same meaning for C_α which x, y, τ have for C. Then if

* See BOLZA, *Lectures on the Calculus of Variations*, p. 117.

† Literal subscripts denote partial differentiation. The expression $f + f_{\tau\tau}$ corresponds to the functions $f_{y'y'}$ and F_1 in other forms of the theory; see BOLZA, loc. cit., pp. 47, 121.

‡ BOLZA, loc. cit., p. 122.

the curve C has the shortest length, the derivative $dl/d\alpha$ must vanish for $\alpha = 0$. This derivative is

$$(4)\quad \left[\frac{dl}{d\alpha}\right]_{\alpha=0} = \int_{t_0}^{t_1} \left\{ \left[f_x \xi + f_y \eta + f_\tau \frac{\partial \bar{\tau}}{\partial \alpha} \right] \sqrt{x'^2 + y'^2} + f \frac{x'\xi + y'\eta}{\sqrt{x'^2 + y'^2}} \right\} dt,$$

where the arguments of f and its derivatives f_x, f_y, f_τ are x, y, τ from equations (1) and (3), and α has the value zero in $\partial \bar{\tau}/\partial \alpha$. The value of the last expression can be calculated from equations (3) in which $\bar{x}$ and $\bar{y}$ must be substituted for x and y:

$$\left[\frac{\partial \bar{\tau}}{\partial \alpha}\right]_{\alpha=0} = \frac{x'\eta' - y'\xi'}{x'^2 + y'^2}.$$

After some simple alterations equation (4) takes the form

$$\left[\frac{dl}{d\alpha}\right]_{\alpha=0} = \int_{s_0}^{s_1} \left\{ f_x \xi + f_y \eta + (f \cos\tau - f_\tau \sin\tau) \frac{d\xi}{ds} + (f \sin\tau + f_\tau \cos\tau) \frac{d\eta}{ds} \right\} ds,$$

in which the variable of integration has been changed into the Euclidean length of arc s. By the usual partial integration of the calculus of variations, one obtains

$$(5)\quad \left[\frac{dl}{d\alpha}\right]_{\alpha=0} = \int_{s_0}^{s_1} \{ P\xi + Q\eta \}\, ds,$$

where

$$P = f_x - \frac{d}{ds}(f \cos\tau - f_\tau \sin\tau), \qquad Q = f_y - \frac{d}{ds}(f \sin\tau + f_\tau \cos\tau).$$

The integral (5) must be zero for all functions ξ and η vanishing at t_0 and t_1 and having continuous derivatives. Hence along the minimizing curve C the two conditions

$$(6)\quad P = 0, \qquad Q = 0$$

must be satisfied. These equations are, however, equivalent to the single equation

$$(7)\quad f_x \sin\tau - f_y \cos\tau + f_{x\tau} \cos\tau + f_{y\tau} \sin\tau + (f + f_{\tau\tau}) \frac{d\tau}{ds} = 0.$$

For denote the left member of the last equation by T. Then the relations

$$P = T \sin\tau, \qquad Q = -T \cos\tau,$$

hold.

The differential equation (7) *is the Euler equation* of the calculus of variations for the integral* (2), *and must be satisfied by every curve C having continuous first and second derivatives which makes the integral I a minimum,*

* BOLZA, loc. cit., pp. 20, 123.

that is, by the shortest lines in the geometry for which the integral (2) *is defined to be the length integral.*

In order to obtain the parametric representation of the extremal, the three equations

$$\frac{dx}{ds} = \cos\tau,$$

$$\frac{dy}{ds} = \cos\tau, \tag{8}$$

$$\frac{d\tau}{ds} = \frac{f_x \sin\tau - f_y \cos\tau + f_{x\tau}\cos\tau + f_{y\tau}\sin\tau}{f + f_{\tau\tau}},$$

must be integrated. From the properties assumed for the function f and the known existence theorems for differential equations,* the following theorem may be stated:

Through a given point x_0, y_0 and direction τ_0 there passes one and but one shortest line (extremal) solving equation (7). *Analytically these extremals have the form*

$$x = \phi(s, x_0, y_0, \tau_0), \qquad y = \psi(s, x_0, y_0, \tau_0), \tag{9}$$

where the initial values of ϕ and ψ at $s = 0$ are such that

$$\begin{aligned} x_0 &\equiv \phi(0, x_0, y_0, \tau_0), & \cos\tau_0 &\equiv \phi_s(0, x_0, y_0, \tau_0), \\ y_0 &\equiv \psi(0, x_0, y_0, \tau_0) & \sin\tau_0 &\equiv \psi_s(0, x_0, y_0, \tau_0). \end{aligned} \tag{10}$$

As for the further properties of the functions ϕ and ψ, the existence theorems show that in a suitably chosen region the solutions x, y, τ of equations (8) must have first partial derivatives continuous in all arguments. From the first of these equations it follows easily that ϕ has also continuous derivatives ϕ_{ss} and $\phi_{s\tau_0}$, and similarly for the function ψ.

As the extremals correspond to straight lines in Euclidean geometry, so another class of curves called in the calculus of variations transversals,† correspond to the Euclidean circles. Consider all the curves (9) which pass through a fixed point $O(x_0, y_0)$. On each of these mark off an arc such that the generalized length has a fixed value r. The value of s determining the end-point A of this arc (see Fig. 1) is defined by the equation

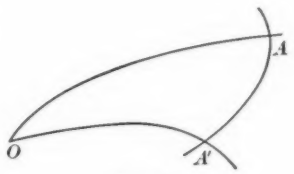


FIG. 1.

$$r = \int_0^s f(\phi, \psi, \tau)\, ds. \tag{11}$$

* BLISS, Annals of Mathematics, vol. 6 (1905), p. 49.

† BOLZA, loc. cit., p. 167.

Since the derivative for s of the right member is $f(x_0, y_0, \tau_0)$ at $s = 0$, and by hypothesis different from zero, this equation can be solved for s in terms of r and τ_0, and the solution will have continuous first partial derivatives for both arguments in a properly chosen region.* Since

$$\frac{\partial s}{\partial r} = \frac{1}{f(\phi, \psi, \tau)}$$

is greater than zero near the point O, the value of s will increase from zero with r. The equation of the transversal $A'A$ is found by substituting this value of s in equations (9). The variables x and y then become functions of τ_0 and r, and the points found by keeping r constant and assigning different values to τ_0 constitute the transversal $A'A$ at the generalized distance r from O.

The slope of the transversal at any point A is not difficult to find. In order that the calculation may be analogous to that made above in deducing EULER's equation, equation (11) may be written

$$r = \int_0^s f(\phi, \psi, \tau)\sqrt{\phi_s^2 + \psi_s^2}\, ds,$$

since

$$\phi_s^2 + \psi_s^2 \equiv 1. \tag{12}$$

Derivation for τ_0 gives

$$0 = f(\phi, \psi, \tau)\frac{\partial s}{\partial \tau_0} + \int_0^s \frac{\partial}{\partial \tau_0}\{f(\phi, \psi, \tau)\sqrt{\phi_s^2 + \psi_s^2}\}\, ds,$$

or by the same process of partial integration used above in finding the extremals,

$$0 = f\frac{\partial s}{\partial \tau_0} + (f\cos\tau - f_\tau \sin\tau)\phi_{\tau_0} + (f\sin\tau + f_\tau\cos\tau)\psi_{\tau_0}. \tag{13}$$

In deriving this result one must remember that the curves (9) are all extremals satisfying equations (6), so that the integral part of the expression disappears. Furthermore ϕ_{τ_0} and ψ_{τ_0} vanish identically when $s = 0$ on account of the initial conditions (10). The value of $dx/d\tau_0$, which with $dy/d\tau_0$ define the slope of the transversal, is found from the equation

$$\frac{dx}{d\tau_0} = \cos\tau\frac{\partial s}{\partial \tau_0} + \phi_{\tau_0},$$

where $\cos\tau$ is substituted in place of its equal ϕ_s. By introducing the value of $\partial s/\partial\tau_0$ from (13) and using the relation (12), it follows that

$$\frac{dx}{d\tau_0} = \frac{\phi_{\tau_0}\sin\tau - \psi_{\tau_0}\cos\tau}{f}(f\sin\tau + f_\tau\cos\tau), \tag{14}$$

* For the theorems on implicit functions, see GOURSAT, *Cours d'Analyse*, p. 40, or the translation of the same by HEDRICK, p. 35.

and by a similar calculation,

$$\frac{dy}{d\tau_0} = \frac{\phi_{\tau_0} \sin\tau - \psi_{\tau_0} \cos\tau}{f}\,(-f\cos\tau + f_\tau \sin\tau). \tag{15}$$

From these two equations it is seen that the direction of the transversal, being determined by the parentheses in the right members, depends only upon the values of x, y, τ at the point A and is independent of the center and radius of the generalized circle.

At any point (x, y) on an extremal whose tangent at (x, y) makes an angle τ with the x-axis, the direction defined by

$$\cos\bar{\tau} = \frac{f\sin\tau + f_\tau\cos\tau}{\sqrt{f^2 + f_\tau^2}}, \qquad \sin\bar{\tau} = \frac{-f\cos\tau + f_\tau\sin\tau}{\sqrt{f^2 + f_\tau^2}},$$

where the arguments of f and f_τ are x, y, τ, is the direction which a generalized circle (transversal) to the extremal through the point (x, y) must have. From the two equations* (14) *and* (15) *the parametric representation of the generalized circle may be found by integration in the form*

$$x = g(\tau_0, r), \qquad y = h(\tau_0, r),$$

where τ_0 is the variable parameter, and r the generalized distance of the circle from the center O.

For in the right hand members of (14) and (15), ϕ, ψ, $\cos\tau = \phi_s$, and $\sin\tau = \psi_s$ are all functions of s and τ_0, and s is to be expressed in terms of r and τ_0 from equation (11).

§ 2. *The notion of angle and its generalization.*

With the preceding results in mind the generalized angle corresponding to the integral (2) may be readily derived. In Euclidean geometry the angle between two straight line segments OA' and OA, is the ratio of the length l of a circular arc $A'A$ about O, divided by the radius r of the circle. This measure of angle is independent of the particular circle chosen, however small its radius may be. In the geometries in which the generalization of length of § 1 is presupposed, the angle between two extremals may be found in an analogous manner. The ratio l/r, in Euclidean geometry independent of the radius, is in general not so, and consequently a limit must be introduced.

If OA' and OA (see Fig. 1) *are two extremals through the point O, and $A'A$ is the arc of length l of a transversal at the generalized distance r from O, then the generalized angle between OA' and OA is defined to be the limit of the ratio l/r as r approaches zero.*

* Compare with BOLZA, loc. cit., pp. 36, 155, 197.

Although the newly defined angle is a somewhat apparent extension of the usual notion, the calculation of its value as a definite integral is not so simple. Let $\bar{x}, \bar{y}, \bar{\tau}$ denote for the present the coördinates and direction of tangent at points on the transversal $A'A$. It will be remembered that $\bar{x}$ and $\bar{y}$ are expressible from (14) and (15) in terms of τ_0 and r, τ_0 being the parameter and r specifying the generalized distance of the transversal from O. If τ_0' and τ_0 are the initial directions of OA' and OA at O, the generalized length of the arc $A'A$ will be

$$l = \int_{\tau_0'}^{\tau_0} f(\bar{x}, \bar{y}, \bar{\tau}) \sqrt{\left(\frac{d\bar{x}}{d\tau_0}\right)^2 + \left(\frac{d\bar{y}}{d\tau_0}\right)^2}\, d\tau_0. \tag{16}$$

The integrand of this integral is a function of τ_0 and r having continuous first partial derivatives, and hence as r approaches zero l approaches as a limit the integral over the integrand taken for $r = 0$. The limit of the ratio l/r depends for its value upon the limit of the integrand in the preceding integral divided by r, which in turn is known when

$$\lim \frac{1}{r}\frac{dx}{d\tau_0} \qquad \text{and} \qquad \lim \frac{1}{r}\frac{dy}{d\tau_0}$$

have been calculated. From equations (14) and (15) it is seen that these last expressions are linear in ϕ_{τ_0}/r and ψ_{τ_0}/r, whose limits can be found in the manner explained in the following paragraph.

By TAYLOR's formula,

$$\phi_{\tau_0}(s, \tau_0) = \phi_{\tau_0}(0, \tau_0) + s\phi_{s\tau_0}(\theta s, \tau_0) \qquad (0 < \theta < 1),$$

where the arguments x_0, y_0 of ϕ, being considered fixed, are omitted. The first term of the right member is identically zero, as has been seen above. From equation (11)

$$\frac{ds}{dr} = \frac{1}{f(\phi, \psi, \tau)},$$

and therefore with the help of equations (10)

$$\lim_{r \doteq 0} \frac{s}{r} = \frac{1}{f(x_0, y_0, \tau_0)}.$$

It follows that

$$\lim_{r \doteq 0} \frac{\phi_{\tau_0}(s, \tau_0)}{r} = \frac{\phi_{s\tau_0}(0, \tau_0)}{f(x_0, y_0, \tau_0)} = -\frac{\sin \tau_0}{f(x_0, y_0, \tau_0)},$$

the last step because, from (10),

$$-\sin \tau_0 \equiv \phi_{s\tau_0}(0, \tau_0).$$

By a similar process of reasoning,

$$\lim \frac{\psi_{\tau_0}(s, r)}{r} = \frac{\cos \tau_0}{f(x_0, y_0, \tau_0)}.$$

From equations (14) and (15),

$$\lim_{r=0} \frac{1}{r}\frac{dx}{d\tau_0} = -\frac{f\sin\tau_0 + f_\tau\cos\tau_0}{f^2},$$
$$\lim_{r=0} \frac{1}{r}\frac{dy}{d\tau_0} = \frac{f\cos\tau_0 - f_\tau\sin\tau_0}{f^2}, \tag{17}$$

where the arguments of f and f_τ are x_0, y_0, τ_0. The limit of l/r follows from equations (16) and (17).

The generalized angle at a point (x, y) between two directions whose Euclidean angles with the x-axis are τ' and τ, has therefore the value

$$\lim_{r=0} \frac{l}{r} = \int_{\tau'}^{\tau} \frac{f(x, y, \bar\tau)}{[f(x, y, \tau)]^2}\sqrt{f^2 + f_\tau^2}\,d\tau, \tag{18}$$

where the arguments of f and f_τ are x, y, τ, and $\bar\tau$ is defined by the equations

$$\cos\bar\tau = \frac{f\sin\tau + f_\tau\cos\tau}{\sqrt{f^2 + f_\tau^2}},$$
$$\sin\bar\tau = \frac{-f\cos\tau + f_\tau\sin\tau}{\sqrt{f^2 + f_\tau^2}}, \tag{19}$$

i. e., *$\bar\tau$ is the direction transverse to τ at the point $O(x, y)$.*

For simplicity the subscript zero designating the point O has been dropped. The integral (18) might be written somewhat more simply

$$\int_{\tau'}^{\tau} \frac{f(x, y, \bar\tau)}{f(x, y, \tau)} \csc(\tau - \bar\tau)\,d\tau.$$

§ 3. *Application to special cases.*

As a first application one would naturally be interested to see the behavior of the integral (18) in the case of ordinary Euclidean geometry. The function $f(x, y, \tau)$ is then simply unity, and the generalized angle becomes $\tau - \tau'$ as would be expected.

The expression for the length of a curve on a surface when the surface is given in parametric representation, is an integral (2) in which the function f has the value

$$f(x, y, \tau) = \sqrt{E\cos^2\tau + 2F\cos\tau\sin\tau + G\sin^2\tau}.* \tag{20}$$

To conform to the previous notations of this paper, x and y are used as the arguments of E, F, and G instead of the customary u, v. As is well known the

*See for example GAUSS, *General Investigations of Curved Surfaces* (translation), p. 20, or his *Werke*, vol. 4, p. 236.

angle ω between any curve and a line $y = \text{const.}$ on the surface, is defined by the formulæ*

$$\cos\omega = \frac{E\cos\tau + F\sin\tau}{\sqrt{E}\sqrt{E\cos^2\tau + 2F\cos\tau\sin\tau + G\sin^2\tau}},$$

$$\sin\omega = \frac{\sqrt{EG - F^2}\cdot\sin\tau}{\sqrt{E}\sqrt{E\cos^2\tau + 2F\cos\tau\sin\tau + G\sin^2\tau}}, \tag{21}$$

where τ now represents the angle between the x-axis and the tangent to the image of the curve in the xy-plane.

The generalized angle defined by the integral (18) *when* $f(x, y, \tau)$ *has the form* (20), *is exactly the same as the ordinary angle* ω *calculated from the formulæ* (21) *between two curves on the surface whose images in the* xy*-plane make angles of* τ' *and* τ *with the* x*-axis.*

In order to prove this it is only necessary to find $d\omega/d\tau$ and show that it has a value identical with that of the integrand of the integral (18) when the value of $f(x, y, \tau)$ from (20) is substituted. The calculation is simplified by the introduction of the notations

$$\lambda = E\cos\tau + F\sin\tau, \qquad \mu = F\cos\tau + G\sin\tau.$$

Then

$$f = \frac{\lambda\cos\tau + \mu\sin\tau}{f}, \qquad f_\tau = \frac{-\lambda\sin\tau + \mu\cos\tau}{f},$$

and from formulæ (19)

$$\cos\bar{\tau} = \frac{\lambda}{\sqrt{\lambda^2 + \mu^2}}, \qquad \sin\bar{\tau} = \frac{-\mu}{\sqrt{\lambda^2 + \mu^2}}.$$

The values of $f(x, y, \bar{\tau})$ and $f^2 + f_\tau^2$ are

$$f(x, y, \bar{\tau}) = \sqrt{EG - F^2}\,\frac{f}{\sqrt{\lambda^2 + \mu^2}}, \qquad f^2 + f_\tau^2 = \frac{\lambda^2 + \mu^2}{f^2},$$

so that the final value of the integrand of (18) is

$$\frac{f(x, y, \bar{\tau})}{f(x, y, \tau)}\csc(\tau - \bar{\tau}) = \frac{\sqrt{EG - F^2}}{E\cos^2\tau + 2F\cos\tau\sin\tau + G\sin^2\tau}, \tag{22}$$

But this is the value of $d\omega/d\tau$, easily calculable from equations (21).

The non-Euclidean geometries in the plane correspond, as is well known, to the geometries on surfaces of constant positive and negative curvature. By a projective transformation of the plane, or of the parameters of the surface according to the point of view which one wishes to take, the length integral can be transformed so that the function $f(x, y, \tau)$ has one of the two forms†

* See GAUSS, Translation, p. 26; *Werke*, vol. 4, p. 242.

† DARBOUX, *Theorie générale des Surfaces*, vol. 3, p. 62; KLEIN, *Nichteuklidische Geometrie*, p. 121.

$$(23)\quad f(x, y, \tau) = \frac{\sqrt{(a^2 \pm y^2)\cos^2\tau \mp 2xy\cos\tau\sin\tau + (a^2 \pm x^2)\sin^2\tau}}{x^2 + y^2 \pm a^2},$$

the upper signs corresponding to elliptic, the lower to hyperbolic geometry. Only the hyperbolic geometry will be considered here on account of the close analogy between the calculations in the two cases.

The elements of the hyperbolic geometry connected with the function (23) are the points interior to the circle

$$(24)\qquad x^2 + y^2 = a^2.$$

It is usual to define the length of a straight line segment OA as a multiple of the logarithm of the anharmonic ratio of the points O and A with the two intersections B_1 and B_2 of the line OA and the circle (24).* It can be shown, how ever, that this definition is identical with the one given above in the form of a definite integral. For denote the coördinates of the point O by (x_0, y_0), of A by (x, y), and let τ be the Euclidean angle which the line OA makes with the x-axis. If s, s_1 and s_2 are the Euclidean distances on the line from the point O to A, B_1 and B_2 respectively, then the hyperbolic length of OA according to the geometrical definition is

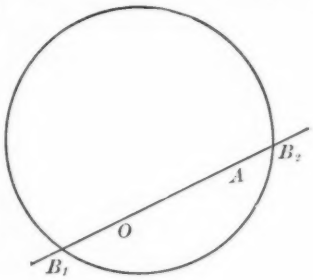


FIG. 2.

$$l = \tfrac{1}{2}\log\left(\frac{s - s_1}{s - s_2}\cdot\frac{s_2}{s_1}\right).$$

The derivative of this expression for s is

$$\frac{dl}{ds} = \tfrac{1}{2}\frac{s_1 - s_2}{(s - s_1)(s - s_2)},$$

which must be equal to the function $f(x, y, \tau)$ defined by equation (23) if the analytical and geometrical definitions of length are to agree in the case of the straight line. The values of $\sigma_1 = s_1 - s$ and $\sigma_2 = s_2 - s$ are easily found by substituting $x + \sigma\cos\tau$ and $y + \sigma\sin\tau$ for x and y in equation (24) and solving for σ. It follows that

$$\sigma_1\sigma_2 = (s - s_1)(s - s_2) = x^2 + y^2 - a^2,$$

$$\sigma_1 - \sigma_2 = s_1 - s_2 = 2\sqrt{(a^2 - y^2)\cos^2\tau + 2xy\cos\tau\sin\tau + (a^2 - x^2)\sin^2\tau},$$

which proves the equality of dl/ds and $f(x, y, \tau)$.

In a similar manner it can be shown that the usual geometrical definition of

* KLEIN, *Nichteuklidische Geometrie*, p. 96.

angle in hyperbolic geometry* gives the same value which one finds from the expression (18). If m_1 and m_2 are the slopes of the two imaginary tangents (dotted in the figure) to the circle (24) through the point $O(x, y)$ interior to the circle, the hyperbolic angle ω between any other two directions making Euclidean angles τ' and τ with the x-axis, is defined geometrically to be a multiple of the logarithm of the anharmonic ratio of the four directions whose tangents are $m_1 = \tan \tau_1$, $m_2 = \tan \tau_2$, $m' = \tan \tau'$, and $m = \tan \tau$; i. e.,

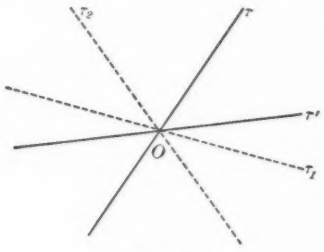


FIG. 3.

$$\omega = \tfrac{1}{2} \log \left(\frac{m - m_1}{m - m_2} \cdot \frac{m' - m_2}{m' - m_1} \right). \tag{25}$$

The derivative of this for τ is

$$\frac{d\omega}{d\tau} = \tfrac{1}{2} \frac{m_1 - m_2}{(m - m_1)(m - m_2)} \sec^2 \tau .$$

But m_1 and m_2 are the roots of the quadratic equation

$$m^2(a^2 - x^2) + 2mxy + (a^2 - y^2) = 0 ,$$

and by substituting one finds

$$\frac{d\omega}{d\tau} = \frac{a \sqrt{a^2 - x^2 - y^2}}{(a^2 - y^2) \cos^2 \tau + 2xy \cos \tau \sin \tau + (a^2 - x^2) \sin^2 \tau} .$$

On the other hand, by comparison of equations (20) and (23),

$$E = \frac{a^2 - y^2}{(a^2 - x^2 - y^2)^2}, \qquad F = \frac{2xy}{(a^2 - x^2 - y^2)^2}, \qquad G = \frac{a^2 - x^2}{(a^2 - x^2 - y^2)^2},$$

and from (22)

$$\frac{f(x, y, \bar{\tau})}{f(x, y, \tau)} \csc (\tau - \bar{\tau}) = \frac{a \sqrt{a^2 - x^2 - y^2}}{(a^2 - y^2) \cos^2 \tau + 2xy \cos \tau \sin \tau + (a^2 - x^2) \sin^2 \tau} .$$

The expression (25), with x, y, τ considered fixed and τ variable, is therefore the primitive of the integral (18).

In the non-Euclidean hyperbolic and elliptic geometries, the geometrical definition of angle as a multiple of the logarithm of the anharmonic ratio of the lines bounding the angle taken with the two tangents to the fundamental conic through their point of intersection, is the same as the analytical definition by means of the definite integral (18).

* KLEIN, loc. cit., p. 87.

In conclusion it might be interesting to indicate a possible generalization of a theorem of Gauss * concerning geodesic triangles. The theorem expresses the excess of the sum of the angles of a geodesic triangle on a surface over two right angles as the double integral of the curvature, multiplied by the element of area, of the surface taken over the same triangle. From this relation one easily derives that the sum of the angles of a triangle is greater than two right angles in elliptic, and less than two right angles in hyperbolic geometry, for these correspond to geometries on surfaces of constant positive and negative curvature respectively. For any problem of the calculus of variations there is an invariant connected with the second variation, more specifically with JACOBI's differential equation, which in the problem of geodesic lines reduces to the curvature of the surface.† It seems certain that a relation similar to GAUSS's must exist between the generalized angles, as defined above, of a triangle whose sides are extremals, and the invariant just mentioned.

THE UNIVERSITY OF CHICAGO,
August, 1905.

* Translation, p. 30; *Werke*, vol. 4, p. 246.
† DARBOUX, loc. cit., p. 94 ff.

THE SQUARE ROOT AND THE RELATIONS OF ORDER*

BY

OSWALD VEBLEN

One of the most obvious discriminations between positive and negative numbers is that the former possess square roots in the field of reals while the latter do not. This distinction, however, has not yet been used in any of the current systems of postulates. On the contrary, an order relation, $<$, is usually introduced as an undefined symbol.

Though the existence of a square root cannot be deduced from order relations without the use of a continuity assumption of some sort, it turns out to be very easy to infer the order relations from postulates about the existence of a square root. Suppose we are given a field, defined by one of the numerous sets of postulates published in the Transactions by HUNTINGTON, MOORE and DICKSON. This field contains a unique mark, 0, such that $x + 0 = x = 0 + x$ for every mark x, and a unique mark, 1, such that $1x = x$ for every mark x. A mark, a, is called a square if there exists a mark, x, such that $xx = a$. If not a square, a mark is called a not-square. Now add the postulates: †

α) *The mark* -1 *is a not-square.*

β) *If marks* x *and* y *are not-squares, then* $x + y$ *is a not-square.*

The postulate α) shows that the marks 2 and 0 are distinct and thus that division by 2 is possible.

Theorem 1. If x and x' are such that $x + x' = 0$, then either x or x' is a square, whereas the other is a not-square.

Proof. If x and x' were both not-squares, then by β), $x + x'$ would be a not-square, whereas zero is a square. The mark $x' = -1 \cdot x$ and hence by the theorem that the product of a square and a not-square is a not-square, if one of the marks x, x' is a square the other is a not square.

Theorem 2. If x and y are squares, then $x + y$ is a square.

* Presented to the Society (Chicago) April 22, 1905. Received for publication November 20, 1905.

† That these postulates are independent is shown by the examples:

α) Any Galois field, $GF(2^n)$.

β) Any Galois field, $GF(p^n)$, where $p > 2$.

Proof. The marks $-x$ and $-y$ are not-squares; hence

$$(-x)+(-y)=-(x+y)$$

is a not-square: and hence, by theorem 1, $x+y$ is a square.

Definition. If and only if the mark $a-b$ is a square which is not zero, $a>b$. If and only if the mark $a-b$ is a not-square, $a<b$.

Theorem 3. If $a \neq b$, either $a>b$ or $a<b$. If $a>b$ then $b<a$; if $a<b$ then $b>a$. If $a<b$ or $b<a$, then $a \neq b$. The inequalities $a>b$ and $a<b$ cannot both be true.

These statements are all evident consequences of the definition and the fact that $a-b$ is either a square or a not-square and cannot be both.

Theorem 4. If $a>b$ and $b>c$, then $a>c$.

Proof. By hypothesis, $a-b$ and $b-c$ are both squares. Hence, by theorem 2, $(a-b)+(b-c)=a-c$ is also a square.

Theorem 5. If $a<b$, then there is a mark, c, such that $a<c$ and $c<b$. Likewise for any mark, a, there are marks d and e such that $a<d$ and $e<a$.

Proof. The mark $d=(b-a)/2$ must, by hypothesis and theorem 2, be a square. Hence $a+d>a$ and $b=a+2d>a+d$. So we have $a+d$ as the mark, c, required by the first part of the theorem. The rest of the theorem is proved in a similar way.

Theorem 6. If $a>0$ and b is any mark, $a+b>b$. If $a<0$ and b is any mark, $a+b<b$.

Proof. In the first case, $a+b-b=a>0$, and in the second case, $a+b-b=a<0$.

Theorem 7. If $a>0$ and $b>0$, then $ab>0$. If $a<0$ and $b>0$, then $ab<0$. If $a<0$ and $b<0$, then $ab>0$.

Proof. These three statements are direct consequences of the propositions that the product of two squares is a square, that the product of a square and a not-square is a not-square and that if b is a square $-1\cdot b$ is a not-square. The corollary is direct that if $a>0$ and $b>1$, $ab>a$. For, writing b in the form $1+e$, $a(1+e)-a=ae$ is evidently a square.

It is now evident, on comparison, for instance, with HUNTINGTON's set of postulates for real algebra on page 39, volume 6 of this journal, that our field satisfies all the conditions of the real number system except that of continuity. We therefore suggest as an elegant way of stating the postulates for the real number system in terms of the undefined symbols, $+$ and $\times$, a set consisting of postulates for a field in general, postulates α) and β), and a continuity postulate. This latter might take the form:

If $[a]$ *and* $[b]$ *are two sets of marks such that* $a-b$ *is for every* a *of* $[a]$ *and* b *of* $[b]$ *a square, not zero, then there exists a mark* d *such that for every* a *of* $[a]$, $a-d$ *is a square and for every* b *of* $[b]$, $d-b$ *is a square.*

What may be thought of as the geometric analogue of the above determination of order relations has been carried out by M. PIERI on page 24 of his paper, *I Principii della Geometria di Posizione*, published in the Memorie dell' Accademia Reale delle Scienze di Torino for 1898. Two point-pairs, AB and CD, are said not to separate or to separate according as there exist or do not exist two points H, H' which are harmonically conjugate both with respect to AB and to CD. Determining the points HH' is equivalent to finding the double points of the involution in which AB and CD are pairs of conjugate points and is therefore a problem of the second degree.

THE PROBLEM OF PARTIAL GEODESIC REPRESENTATION*

BY

EDWARD KASNER

§ 1. *The general problem.*

The original problem of geodesic representation, solved by DINI (1869) with reality restrictions and by LIE (1883) in complete generality, may be stated as follows: Find all pairs of surfaces S and S_1 whose points may be put into correspondence in such a way that every geodesic on the one surface is pictured by a geodesic on the other. Apart from the trivial case where S_1 is isometric with (applicable on) S, or with a surface homothetic to S, the only pairs of surfaces obtained belong to the LIOUVILLE class

$$ds^2 = (U + V)(du^2 + dv^2),$$

or to the LIE class

$$ds^2 = (u + V)\,du\,dv,$$

the latter arising only with imaginary representations.

In this paper the problem is extended by requiring *some*, instead of *all*, the geodesics to correspond. The new problem is made precise by the following considerations:

In the first place, if only two simply infinite systems† of geodesics, on each surface, are to correspond, no restriction is thereby imposed on the pair of surfaces. In fact if S and S_1 are taken arbitrarily, we may select any two systems of geodesics on each surface; if now we take these as parameter lines $u = \text{const.}$, $v = \text{const.}$, and let corresponding points be those represented by the same values of (u, v), our requirement is satisfied.

On the other hand we have the

THEOREM. *If in a point to point correspondence between two surfaces four simply infinite systems of geodesics on each surface correspond, then all geodesics necessarily correspond.*

For the proof we suppose here, as in the rest of the paper, that the gaussian parameters on the two surfaces are selected so that corresponding points are represented by the same values of (u, v), and let the length elements be

* Presented to the Society, October 31, 1903; the results in § 4 were presented separately August 31, 1903. Received for publication November 23, 1905.

† By a simply infinite system we mean one such that through each point of the region on the surface considered there passes one curve of the system and only one.

$$(1)\qquad \begin{aligned} ds^2 &= E du^2 + 2F du\, dv + G dv^2, \\ ds_1^2 &= E_1 du^2 + 2F_1 du\, dv + G_1 dv^2. \end{aligned}$$

The geodesics are defined by differential equations of the form

$$(2)\qquad \begin{aligned} v'' &= Av'^3 + Bv'^2 + Cv' + D, \\ v'' &= A_1 v'^3 + B_1 v'^2 + C_1 v' + D_1. \end{aligned}$$

For corresponding geodesics, if any exist, these equations hold simultaneously, and therefore

$$(3)\qquad (A - A_1)v'^3 + (B - B_1)v'^2 + (C - C_1)v' + (D - D_1) = 0.$$

According to our hypothesis, this is satisfied by four different values of v'; hence it must hold identically. This means that equations (2) coincide; hence all geodesics correspond and our theorem is proved.

The following corollaries deserve explicit statement:

If a surface can be mapped upon the plane in such a way that four systems of geodesics are pictured by straight lines, then all geodesics are so pictured, and hence the surface has constant curvature.*

If a point transformation of the plane converts four systems of straight lines into straight lines, then all straight lines are so converted, and the transformation is thus a collineation.†

Our problem of "partial" geodesic representation can now be stated definitely: *Determine all cases in which two surfaces correspond point by point so that there exist precisely three simply infinite systems of corresponding geodesics.*

That this problem is of much greater generality than the DINI problem is seen as follows. If three systems are to correspond, then equation (3) above must represent geodesics on each surface. If we express the fact that (3) satisfies say the first of equations (2) (the second is then necessarily satisfied also), we are led to a set of three partial differential equations of the second order in E, F, G, E_1, F_1, G_1. To satisfy these, E, F, G may be taken at

* A special form of this proposition was obtained by FINSTERWALDER in his interesting report on the mechanics of surface deformation published in the Jahresberichte der deutschen Mathematiker-Vereinigung, vol. 6 (1899), p. 50: If a surface contains four "linear systems" of geodesics, it must have constant curvature. This means here that there exist four equations of the form $a_i u + b_i v = \text{const.}$ $(i = 1, 2, 3, 4)$ which represent geodesics. This is equivalent to the requirement that four systems of geodesics shall be pictured in the plane by four systems of *parallel* straight lines. The result is true, however, for *any* systems of straight lines. The special form is given in LILIENTHAL, Encyklopädie der Mathematischen Wissenschaften, vol. 3 (1903). p. 351; and STÄCKEL, Mathematische Annalen, vol. 56 (1902), p. 502. In this paper Stäckel proposes the problem of determining all surfaces containing three "linear systems" of geodesics and finds certain solutions.

† This result was given by the author in a note on *The characterization of collineations*, Bulletin American Mathematical Society, vol. 9 (1903), p. 545.

random, and E_1, F_1, G_1 will still contain an infinite number of arbitrary constants. These constants may be selected so that (3) does not reduce to an identity, i. e., so that the surface S_1 is not merely isometric with a surface homothetic to S. Hence for any surface S there exist an infinity of distinct surfaces S_1 satisfying the conditions of our problem.*

§ 2. *The conformal problem.*

If the representation of S_1 on S is conformal, the minimal lines correspond. This accounts for two simple systems of geodesics. The theorem of § 1 then takes the form:

If a surface S_1 can be represented conformally on a surface S in such a way that two real systems of geodesics, on each surface, correspond, then all geodesics correspond. By known theorems the surfaces must then, except for homothetic transformation, be isometric.

Our problem now is to determine all pairs of surfaces which admit of conformal representation with one real system of corresponding geodesics.

Since the representation is conformal, the length elements of both surfaces may be assumed in the isothermal form

$$ds^2 = e^{2\omega}(du^2 + dv^2), \qquad ds_1^2 = e^{2\omega_1}(du^2 + dv^2). \tag{4}$$

The geodesics are defined by

$$v'' = (\omega_v - v'\omega_u)(1 + v'^2), \qquad v'' = (\omega_{1v} - v'\omega_{1u})(1 + v'^2). \tag{5}$$

Geodesics on either surface which correspond to geodesics on the other must satisfy both these equations; hence, disregarding the factor $1 + v'^2$ leading to minimal lines, we find

$$\delta_v - v'\delta_u = 0, \tag{6}$$

where

$$\delta = \omega - \omega_1. \tag{7}$$

The condition that (6) shall define geodesics is found to be

$$\delta_u\delta_v(\delta_{uu} - \delta_{vv}) - \delta_{uv}(\delta_u^2 - \delta_v^2) = (\omega_u\omega_{1v} - \omega_v\omega_{1u})(\delta_u^2 + \delta_v^2), \tag{8}$$

which, as should be the case, is unaffected by the interchange of ω and ω_1. It is easy to verify that (8) is also the condition that the curves $\delta = \text{const.}$ on either S or S_1 shall be (geodesically) parallel. When this is the case the system of geodesics (6) consists of the orthogonal trajectories of these parallels.

* A given element $ds^2 = Edu^2 + 2Fdudv + Gdv^2$, of course, defines not one surface but a class of isometric surfaces; it is, however, unnecessary to distinguish between the members of such a class since our problem deals with the internal geometry of surfaces.

Hence all solutions S, S_1, of the conformal problem may be obtained as follows: Let S be any surface assumed given in the isothermal form

$$ds^2 = e^{2\omega}(du^2 + dv^2);$$

then the related surfaces S_1 are of the form

$$ds_1^2 = e^{2(\omega+\delta)}(du^2 + dv^2),$$

where δ is any function of u, v, such that the curves $\delta = const.$ on S are parallel. By the theory of parallels, δ is therefore an arbitrary function of ϕ, where ϕ is the general solution of

$$\phi_u^2 + \phi_v^2 = e^{2\omega}. \tag{9}$$

If the finite equation of the geodesics on S is known, then all related surfaces S_1 may be found by quadratures. This follows from the known theorem that the determination of the orthogonal trajectories of a given system of geodesics depends upon quadratures.

In particular, the result applies to the determination of surfaces which can be mapped conformally on the plane so that a system of geodesics is pictured by straight lines. All such surfaces are given by

$$ds^2 = e^{2\omega}(du^2 + dv^2),$$

where the function ω is such that the curves $\omega = \text{const.}$ in the u, v plane are parallel. The explicit result is that ω is an arbitrary function f of the argument

$$\phi = \int \frac{dv - t\,du}{\sqrt{1+t^2}},$$

where t is defined by any equation of the form

$$u + vt = \psi(t).$$

Thus ω involves two arbitrary functions f and ψ.

§ 3. *Geodesic parameters.*

Another analytical representation of the surface S, S_1, which is more convenient for the problem discussed in § 4, may be obtained by using gaussian geodesic parameters. On each of the surfaces let the assumed system of geodesics be $v = \text{const.}$, and let their orthogonal trajectories, which also correspond since the representation is conformal, be $u = \text{const.}$ Then the length elements take the form

$$ds^2 = E(u)\,du^2 + G(u, v)\,dv^2,$$

$$ds_1^2 = E_1(u)\,du^2 + G_1(u, v)\,dv^2.$$

The condition for conformality is

$$\frac{G_1}{G} = \frac{E_1}{E}.$$

By a proper selection of the parameter u the function E may be reduced to unity. We may then write

$$\begin{aligned} ds^2 &= du^2 + G dv^2, \\ ds_1^2 &= U(du^2 + G dv^2), \end{aligned} \tag{10}$$

where G is any function of u, v and U any function of u.

All pairs of surfaces S, S_1 which admit of conformal representation with a real system of corresponding geodesics may be reduced to the form (10). *The converse also holds.*

If any surface S is given with an assigned system of geodesics on it, there exist an infinite number of related surfaces S_1 depending upon an arbitrary function U.

In any conformal representation the ratio of corresponding length elements, $\rho = ds_1/ds$, is independent of the direction of those elements and is thus a point function. In our case we have, from (10),

$$\frac{ds_1}{ds} = \sqrt{U}.$$

Since the curves $u = \text{const.}$ are parallel, we have the result:

In a conformal representation of the kind considered, the curves, on either surface, for which the ratio ds_1/ds is constant are geodesically parallel.

The converse, which may be proved without much difficulty, takes the form:

If in any conformal representation the curves $ds_1/ds = \text{const.}$ are parallel on one surface, they are also parallel on the other surface; the geodesics orthogonal to these parallels then correspond by the representation.

We now determine the surfaces S_1 which can be represented on a plane. The element ds^2 in (10) must have zero curvature. From this we find that G must have the form

$$G = (V_1 u + V)^2,$$

where V, V_1 are any functions of v.

Therefore all cases of surfaces S_1 which can be conformally represented on a plane S so that ∞^1 geodesics are pictured by straight lines are given by

$$\begin{aligned} ds^2 &= du^2 + (V_1 u + V)^2 dv^2, \\ ds_1^2 &= U\{du^2 + (V_1 u + V)^2 dv^2\}. \end{aligned} \tag{11}$$

If $V_1 = 0$ and $V = 1$, the element ds_1 belongs to a surface of revolution. Thus the surfaces applicable on surfaces of revolution enter as particular solutions of our problem.

§ 4. *A class of conformal transformations in the plane.*

The case in which both S and S_1 are planes is especially interesting. The question is then to determine all conformal representations of a plane on itself (or on another plane) by which a simply infinite system of straight lines are pictured by straight lines. By the theorem of § 1, there cannot be two real systems unless the representation reduces merely to a conformal collineation, i. e., a similitude transformation.

All cases where S has zero curvature are included in (11); the question is then to determine the functions U, V, V_1, so that S_1 shall have zero curvature also.

We divide the discussion according to the vanishing or non-vanishing of V_1.

If $V_1 = 0$, the elements become, after a proper change in the parameter v and the substitution of U^2 for U,

$$ds^2 = du^2 + dv^2, \qquad ds_1^2 = U^2(du^2 + dv^2). \tag{12}$$

The condition that the curvature of ds_1 shall vanish is

$$UU'' - U'^2 = 0;$$

hence

$$U = e^{au+b}.$$

The solution obtained is

$$ds^2 = du^2 + dv^2, \qquad ds_1^2 = e^{2(au+b)}(du^2 + dv^2). \tag{13}$$

To find the corresponding conformal transformations, we reduce (13) to the minimal form by putting

$$d\alpha = du + idv, \qquad d\beta = du - idv.$$

This gives

$$ds^2 = d\alpha d\beta, \qquad ds_1^2 = e^{a(\alpha+\beta)+b}\, d\alpha d\beta. \tag{14}$$

Now every conformal transformation is equivalent to a substitution of some function $f(\alpha)$ for α and some function $g(\beta)$ for β. The condition that ds^2 is converted into ds_1^2 is

$$f'(\alpha)g'(\beta) = e^{a(\alpha+\beta)+b}.$$

The solution of this functional equation is

$$f = e^{c_1\alpha+a_1} + a_2, \qquad g = e^{c_1\beta+b_1} + b_2. \tag{15}$$

The second case, where V_1 is not zero, is more complicated. The elements (11) may now be reduced to

$$ds^2 = du^2 + (u+V)^2 dv^2, \qquad ds_1^2 = U^2\{du^2 + (u+V)^2 dv^2\}.$$

The condition that S_1 has zero curvature is

$$UU' + (u+V)(UU'' - U'^2) = 0. \tag{16}$$

The case where $UU'' - U'^2$ vanishes is trivial since (16) shows that then U must be a constant; then we have merely a similitude transformation.

We may therefore write (16) in the form

$$\frac{UU'}{UU'' - U'^2} + u = -V. \tag{17}$$

Each member of this equation must reduce to the same constant; this may without loss of generality be assumed zero. We find

$$V = 0, \qquad U = au^c$$

so that the new solution is

$$ds = du^2 + u^2 dv^2, \qquad ds_1^2 = a^2 u^{2c}(du^2 + u^2 dv^2). \tag{18}$$

The reduction to minimal parameters, by means of the substitution

$$d\alpha = e^{iv}(du + iudv), \qquad d\beta = e^{-iv}(du - iudv),$$

gives

$$ds^2 = d\alpha d\beta, \qquad ds_1^2 = a\alpha^c \beta^c d\alpha d\beta. \tag{19}$$

The equation for the determination of f and g is now

$$f'(\alpha) g'(\beta) = a\alpha^c \beta^c.$$

The solutions are found to be of these two types:

$$f = a_1 \alpha^{c+1} + a_2, \qquad g = b_1 \beta^{c+1} + b_2; \tag{20}$$

$$f = a_1 \log \alpha + a_2, \qquad g = b_1 \log \beta + b_2. \tag{21}$$

We thus find that all conformal transformations of the kind required are given by (15), (20) and (21). For real transformations f and g must be conjugate, so that it is sufficient to consider say f. Using the usual complex variable z instead of α, we find that the real transformations are

$$Z = e^{az+b}, \qquad Z = a \log z + b, \qquad Z = az^k + b, \tag{22}$$

where a, b are any (real or complex) constants and k is real. By means of similitude transformation these may be reduced to the canonical forms

$$Z = e^z, \qquad Z = \log z, \qquad Z = z^k. \tag{23}$$

Every real conformal transformation of the plane into itself which converts a single infinity of real straight lines into straight lines belongs to one of the three types (23). *In the exponential type a set of parallels is converted into a pencil (with a finite vertex); in the logarithmic type a pencil is converted into a set of parallels; in the monomial type a pencil is converted into a pencil.*

The first two types may be regarded as limiting cases of the last type.

COLUMBIA UNIVERSITY.

ON THE PENTADELTOID*

BY

R. P. STEPHENS

INTRODUCTORY.

In his *Orthocentric Properties of the Plane n-Line*, Professor MORLEY† makes use of a curve of class $2n-1$, which he calls the Δ^{2n-1}. Of these curves, the deltoid ($n=2$) is the well-known three-cusped hypocycloid; but the curve of class five ($n=3$), which I shall call *Pentadeltoid*, and more frequently represent by the symbol Δ^5, is not so well known, though its shape and a few of its properties were known to CLIFFORD‡. I shall here prove some of the properties of the pentadeltoid. My purpose in presenting this curve is twofold: 1. the discussion of a curve of many interesting properties; 2. the illustration of the use of vector analysis as a tool in the metric geometry of the plane.

The line equation of a curve is here uniformly expressed by means of conjugate coördinates,§ while the point equation is expressed by means of what is commonly called the map equation, that is, a point of the curve is expressed as a function of a parameter t, which is limited to the unit circle and is, therefore, always equal in absolute value to unity. Throughout I shall consider y and b as conjugates of x and a, respectively; that is, if X and Y are rectangular coördinates of a point, then

$$x = X + iY,$$

$$y = X - iY,$$

and similarly for a and b.‖

Some of the properties of the curves of class $2n-1$ and of degree $2n$, easily inferred from those of the Δ^5, are also proved, but there has been no attempt at making a complete study of the general case.

§ 1. *The construction of the curve.*

The line equation of the pentadeltoid, as given in Professor MORLEY's article referred to above, is of the form

* Presented to the Society December 30, 1905. Received for publication September 15, 1905.

† Transactions of the American Mathematical Society, vol. 4 (1903), pp. 1–12.

‡ W. K. CLIFFORD: *Mathematical Papers*, pp. 614–617, and plate opposite p. 642.

§ F. FRANKLIN: *Some applications of circular coördinates*, American Journal of Mathematics, vol. 12 (1890), p. 161. F. MORLEY: *On the Metric Geometry of the Plane n-Line*, Transactions of the American Mathematical Society, vol. 1 (1900), p. 97.

‖ H. A. CONVERSE: Annals of Mathematics, ser. 2, vol. 5 (1904), pp. 106–109, where, in a paper closely allied to this, is given a fuller explanation of these methods.

$$t^5 + at^4 + xt^3 + yt^2 + bt + 1 = 0,$$

where t is a turn and a and b are conjugates. For a fixed value of t, this is the equation of a line; but, as t describes the unit circle, this line envelopes the Δ^5.

The equation may be written in the form

$$tx + y = t\left(-t^2 - \frac{b}{t^2}\right) + \left(-\frac{1}{t^2} - at^2\right),$$

whence it is evident that it is a line with clinant $-1/t$ through the point

$$x_1 = -t^2 - \frac{b}{t^2}.$$

For varying t, x_1 describes an ellipse with center at the origin; hence the Δ^5 is the envelope of a line with clinant $-1/t$ which passes through the point

$$x_1 = -t^2 - \frac{b}{t^2}$$

as t varies. Or, putting it more concretely, we may say that if a straight edge be attached to the generating point of an ellipse and given the proper rotation about this point, its envelope will be a Δ^5.

The ellipse can best be generated by means of the instrument commonly called the ellipsograph. The desired rotation of the straight edge can be secured by cog-wheels, properly geared. In the figure $1A$, suppose the segment AB of length 2 be allowed to move with its extremities always in the perpendicular lines AO and OB, then the point M, the mid-point of AB, will trace out a unit circle. At any instant in its motion in a positive direction about O, let M be at τ^2; then, if MP in absolute value is μ, we have for the point P, since it revolves about M in a negative direction at the same rate as M about O,

$$x_1 = \tau^2 + \frac{\mu}{\tau^2}.$$

Centered at A on both AB and AO is a cog-wheel C_3, which is so attached that its only motion is horizontal translation along AO. Centered on AB is another cog-wheel C_4, which moves with AB, but also may rotate about P as a center. These two cog-wheels, C_3 and C_4, whose diameters are in the ratio of 3 to 4, are connected by the small cog-wheel C. The line l passing through P rotates with C_4.

As OM moves positively through an angle θ (a turn τ^2), the cog-wheel C_4, owing to the motion of AB, rotates with reference to the fixed plane through a negative angle $-\theta$; but, on account of its connection with C_3, it is turned in a positive direction $\frac{3}{4}\theta$, which leaves it with a resultant rotation $-\frac{1}{4}\theta$, or a turn

$1/1\sqrt{\tau}$. Hence, if the clinant of l was initially $-\kappa$, it is now $-\kappa/\tau$. Therefore, since l goes through the point

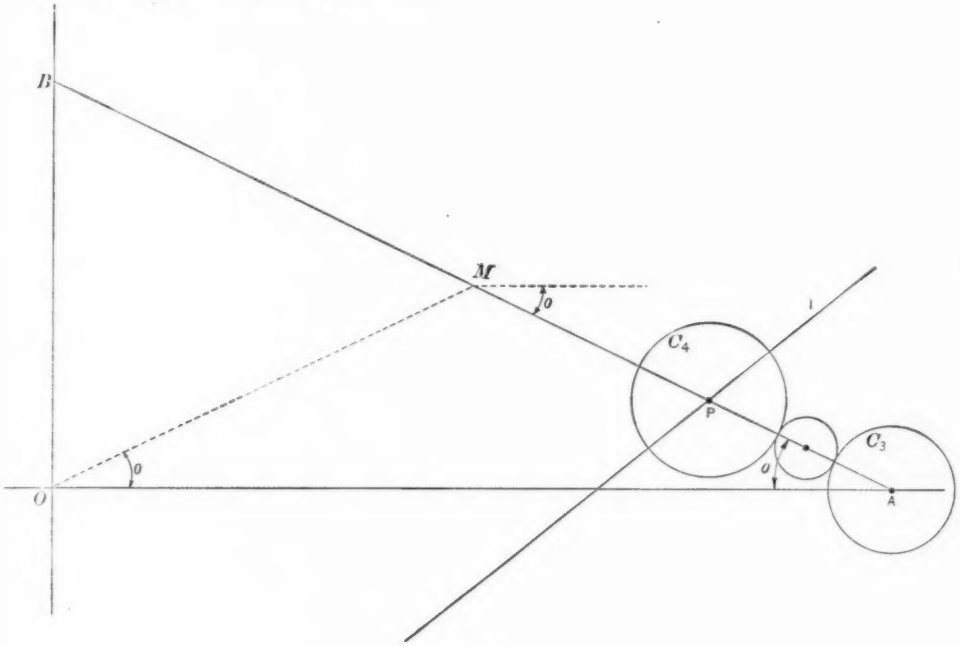


FIG. 1A.

$$x_1 = \tau^2 + \frac{\mu}{\tau^2},$$

its equation is

$$x + \frac{\kappa}{\tau} y = \left(\tau^2 + \frac{\mu}{\tau^2}\right) + \frac{\kappa}{\tau}\left(\frac{1}{\tau^2} + \mu\tau^2\right);$$

or putting

$$it \equiv \tau \qquad \text{and} \qquad i\alpha \equiv \kappa,$$

we have, as the equation of the line l,

$$t^5 + \alpha\mu t^4 + xt^3 + \alpha y t^2 + \mu t + \alpha = 0, \tag{1}$$

where t and α are *turns* and μ is *real* and *positive*. The two constants here, α and μ, are of the same effect as the two conjugates, a and b, used above, so this new equation is just as general as the first form.

If, in the mechanism described above, a cog-wheel C_5 be used for C_3, that is, if cog-wheels whose diameters have the ratio 5 to 4 instead of 3 to 4 be used, the equation of the curve will have the form

$$\mu t^5 + \alpha t^4 + xt^3 + \alpha y t^2 + t + \alpha\mu = 0.$$

The advantage of this double description is easily seen. If $\mu = 0$, then the first form becomes

$$t^5 + xt^3 + \alpha y t^2 + \alpha = 0,$$

which, as t varies, envelopes the curve called by Clifford the *five-rayed star*, but which I shall call a regular Δ^5. However, if $\mu = 0$ in the second form, we have

$$\alpha t^3 + xt^2 + \alpha y t + 1 = 0,$$

which is the line equation of the deltoid. So with the two combinations of cog-wheels, which can be adjusted in the same mechanism, we can get all shapes of the Δ^5 as it varies from the five-rayed star to the deltoid.

The diagram of Fig. 1*B* shows the instrument with which the curves in the other figures were described.

§ 2. *The equation of the curve.*

The Δ^5 as here treated has the equation

$$t^5 + \alpha\mu t^4 + xt^3 + \alpha y t^2 + \mu t + \alpha = 0. \tag{1}$$

If we divide this equation by t^2 and then differentiate as to t, we shall obtain the map equation of the curve. Thus

$$x = -\left(3t^2 + 2\alpha\mu t - \frac{\mu}{t^2} - \frac{2\alpha}{t^3}\right), \tag{2}$$

where t is the variable turn. Evidently for any given value of t, (2) is that point of the curve at which the line (1) is tangent.

From equation (1) it is seen that from any point x there are five tangents to the curve; that is, it is of the *fifth class*. By substituting x and its conjugate from equation (2) in the equation of a line, we obtain a sextic in t; hence the curve is of the *sixth order*.

Again, since the sum of five turns can never be greater in absolute value than 5, it is seen from (1), that if $\mu > 5$, there is no point of the plane from which five tangents, all of them real, can be drawn to the curve.

Whenever reference is made to the general case,* the equation of the curve of class $4n - 1$ will be assumed of the form:

$$\alpha t^{4n-1} + a_1 t^{4n-2} + \alpha a_2 t^{4n-3} + \cdots + \alpha\mu t^{2n+1} + xt^{2n}$$
$$+ \alpha y t^{2n-1} + \mu t^{2n-2} + \cdots + b_2 t^2 + \alpha b_1 t + 1 = 0;$$

* F. Morley: Transactions, loc. cit.

and of class $4n+1$:

$$t^{4n+1} + \alpha a_1 t^{4n} + a_2 t^{4n-1} + \cdots + \alpha\mu t^{2n+2} + \alpha t^{2n+1}$$
$$+ \alpha y t^{2n} + \mu t^{2n-1} + \cdots + \alpha b_2 t^2 + b_1 t + \alpha = 0,$$

where α and t are turns, μ is real, and a_i and b_i are conjugates. These forms

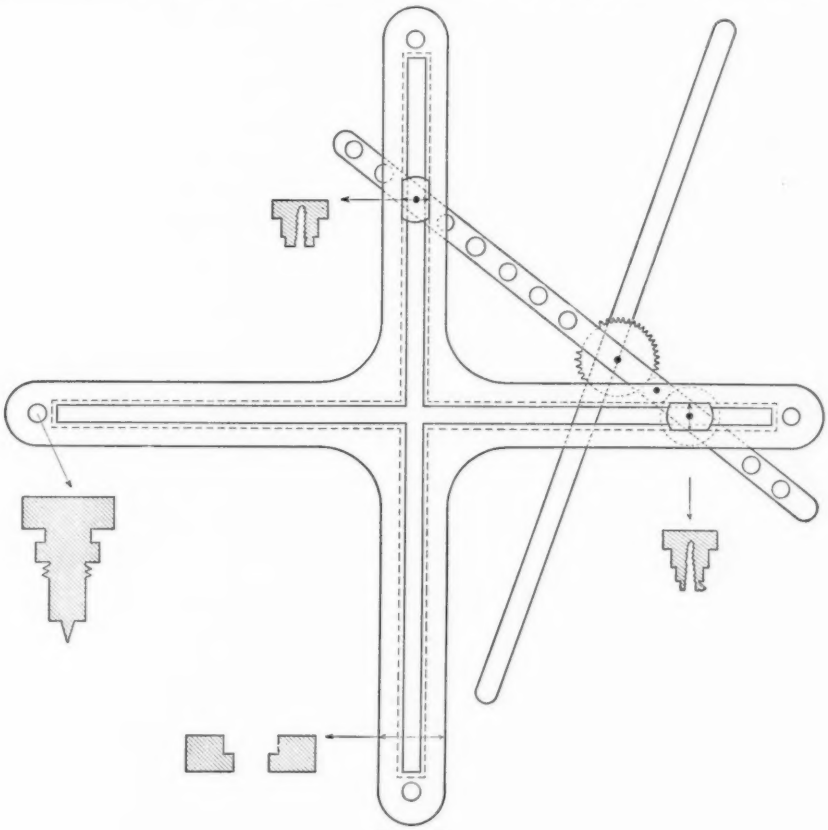


FIG. 1*B*.

are derived in a manner analogous to the case of the Δ^5. The curve Δ^7, for instance,

$$t^7 + a_1 t^6 + a_2 t^5 + xt^4 + yt^3 + b_2 t^2 + b_1 t + 1 = 0,$$

can be written in the form

$$tx + y = -t\left(a_1 t^2 + \frac{b_2}{t^2} + \frac{1}{t^4}\right) - \left(\frac{b_1}{t^2} + a_2 t^2 + t^4\right),$$

obviously a line through the point

$$x = -\left(a_1 t^2 + \frac{b_2}{t^2} + \frac{1}{t^4}\right).$$

By a proper choice of a unit circle, one of the two complex constants, a_1 and b_2, can be made real, then we have

$$x = -\left(at^2 + \frac{\mu}{t^2} + \frac{1}{t^4}\right),$$

where μ is real. The Δ^7 is, therefore, the envelope of a line through this point as t varies, provided the line be given the proper rotation about this point. Owing to our change in the unit circle, we must take account of the initial inclination of the line and so must introduce a constant turn, α. Therefore the equation of the line is

$$tx + \alpha y = -t\left(at^2 + \frac{\mu}{t^2} + \frac{1}{t^4}\right) - \alpha\left(\frac{b}{t^2} + \mu t^2 + t^4\right).$$

This may be put into the form

$$\alpha t^7 + at^6 + \alpha\mu t^5 + xt^4 + \alpha y t^3 + \mu t^2 + bt + \alpha = 0,$$

a line which for varying t envelopes the Δ^7.

So in general we can replace one of the complex constants of the Δ^{2n-1}, when in the form given by Professor MORLEY, by a real number and a turn, and thus obtain the general equations as given above.

§ 3. *Perpendicular tangents.*

If equation (1) is a tangent to the curve, then obviously the tangent perpendicular to it is obtained by putting $-t$ for t in (1). Thus,

$$-t^5 + \alpha\mu t^4 - xt^3 + \alpha y t^2 - \mu t + \alpha = 0.$$

The point of intersection of these two lines is

$$x = -\left(t^2 + \frac{\mu}{t^2}\right), \tag{3}$$

which for varying t is an ellipse, in fact the same ellipse which appeared above in the construction of the curve. Hence, *perpendicular tangents of a Δ^5 intersect along an ellipse.*

This ellipse, the orthoptic curve, has its center at the origin and its foci on the axis of reals. For $\mu = 0$, it becomes the unit base circle and for $\mu = 1$, it becomes the segment of a line. For convenience, I shall refer to the center of the orthoptic curve of the Δ^5 as the center of that Δ^5.

That this ellipse is tangent to the Δ^5 can be shown in this way. The clinant of the tangent to the ellipse at any point is

$$\frac{dx}{dy} = \frac{\dfrac{dx}{dt}}{\dfrac{dy}{dt}} = \frac{t^4 - \mu}{\mu t^4 - 1}.$$

Evidently, one of the two perpendicular tangents to the Δ^5 which intersect on the ellipse will be tangent to the ellipse at that point of intersection, provided its clinant equals the clinant above, that is, if

$$(4) \qquad \frac{t^4 - \mu}{\mu t^4 - 1} = -\frac{\alpha}{t}, \qquad \text{or} \qquad t^5 + \alpha\mu t^4 - \mu t - \alpha = 0.$$

There are therefore five common tangents of the Δ^5 and its orthoptic ellipse.

Now the points at which a common tangent touches the two curves have the same parameter t, and hence coincide if

$$t^2 + \frac{\mu}{t^2} = 3t^2 + 2\alpha\mu t - \frac{\mu}{t^2} - \frac{2\alpha}{t^3}.$$

This reduces to

$$t^5 + \alpha\mu t^4 - \mu t - \alpha = 0,$$

which is condition (4) above; so a common tangent touches the two curves in coincident points. Therefore, *the Δ^5 touches its orthoptic ellipse in five points.* (See Fig. 2.)

In an exactly similar manner, the corresponding theorem may be proved for the Δ^{2n-1}, that is, *the Δ^{2n-1} is tangent to its orthoptic curve in $2n-1$ points.*

From (4), it is evident that at least two of the common tangents become imaginary for $\mu > 5$. This is a sufficient but not a necessary condition for imaginary common tangents.

The general contact theorem above could have been stated otherwise, thus: *the Δ^{2n-1} has $2n-1$ tangents which are also normals not only to itself but also to its orthoptic curve.* The equation giving the parameters of these normals in the case of the Δ^5 is the result of substituting $-t$ for t in equation (4); it is therefore

$$(5) \qquad t^5 - \alpha\mu t^4 - \mu t + \alpha = 0.$$

If we subtract this from equation (1), the line equation of the Δ^5, we obtain

$$2\alpha\mu t^3 + xt^2 + \alpha yt + 2\mu = 0,$$

the line equation of a deltoid whose map equation is

$$x = -2\mu\left(2\alpha t - \frac{1}{t^2}\right).$$

From this it is seen that *the five tangents of a* Δ^5 *which are also normals to itself and its orthoptic ellipse touch a concentric deltoid.*

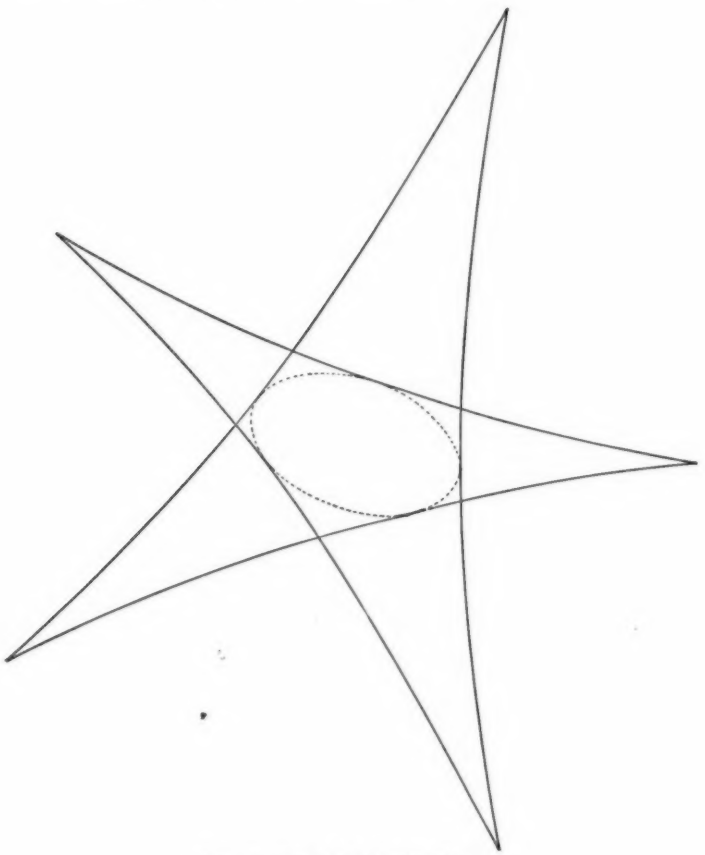


FIG. 2. Pentadeltoid when $\frac{\alpha}{\beta} = .2$.

In a similar manner, *the* $2n-1$ *tangents of a* Δ^{2n-1} *which are normals common to itself and its orthoptic curve touch a concentric* Δ^{2n-3}.

If equation (5) be added to (1), then

$$2t^5 + xt^3 + \alpha y t^2 + 2\alpha = 0,$$

or

$$x = -2\left(3t^2 - \frac{2\alpha}{t^3}\right),$$

the equation of a regular Δ^5, which is independent of μ. Hence, *the five tan-*

gents of a Δ^5 which are normals common to itself and its orthoptic ellipse touch a concentric regular Δ^5; and such normals, for a system of Δ^5's having the same α, envelope the same regular Δ^5 for a varying μ.

There is an interesting generalization suggested by the numbers required to determine the orthoptic curve of the deltoid and the pentadeltoid. In each case the number of points of contact of the orthoptic curve and the Δ^3 or Δ^5 is just sufficient to determine the orthoptic curve. That this law holds in general is easily proved. Consider a curve of class $4n-1$ given by

$$\alpha t^{4n-1} + a_1 t^{4n-2} + \alpha a_2 t^{4n-3} + \cdots + \alpha\mu t^{2n+1} + xt^{2n} + \alpha y t^{2n-1} + \cdots + \alpha b_1 t + 1 = 0.$$

Its orthoptic curve is obtained just as was that of the Δ^5. It is

$$x = -\left(a_1\tau^{n-1} + a_3\tau^{n-2} + \cdots + a_{2n-3}\tau + \frac{\mu}{\tau} + \cdots + \frac{1}{\tau^n}\right),$$

where $\tau \equiv t^2$. This is a curve of order $2n$. In counting its constants we must remember that the base circle requires four constants and that each complex number carries with it its conjugate. Hence, the numbers for the orthoptic curve are:

Base circle		4 constants,
Constants	$a_1, a_3, a_5, \cdots, a_{2n-3}$, $2\left[\frac{2n-3+1}{2}\right] = 2n-2$	" ,
Constants	$b_2, b_4, \cdots, b_{2n-4}$, $2\left[\frac{2n-4}{2}\right] = 2n-4$	" ,
Constant	μ,	1 " ,
Total		$4n-1$ constants.

But we have seen that $4n-1$ is the number of points of contact of the Δ^{4n-1} and its orthoptic curve. In a similar manner, the curve,

$$t^{4n+1} + \alpha a_1 t^{4n} + a_2 t^{4n-1} + \cdots + \alpha\mu t^{2n+2} + xt^{2n+1} + \alpha y t^{2n} + \mu t^{2n-1} + \cdots + b_1 t + \alpha = 0,$$

has for its orthoptic curve

$$x = -\left(\tau^n + a_2\tau^{n-1} + \cdots + a_{2n-2}\tau + \frac{\mu}{\tau} + \cdots + \frac{b_1}{\tau^n}\right),$$

a curve of order $2n$. Counting the constants here as in the other case, we find $4n+1$, which again agrees with the number of points of contact. Therefore, in general, *the points of tangency of a Δ^{2n-1} with its orthoptic curve are just sufficient to determine that curve.*

The degree of the orthoptic curve of the Δ^{4n-1} is the same as that of the orthoptic curve of the Δ^{4n+1} — each is $2n$. But in all cases the numbers

necessary to determine the Δ^{2n-1} are just one more than those required for its orthoptic curve. In the equations given this extra number is α. From this comes an important theorem: *Connected with the orthoptic curve of any Δ^{2n-1} is a singly infinite system of Δ^{2n-1}'s, all having the same orthoptic curve.*

For the Δ^5, this last theorem is seen directly. The inscribed ellipse (3) is obviously independent of α, and hence *there is a single infinity of Δ^5's, each of which touches the same ellipse in five points.*

§ 4. *Nodes and their locus.*

In a previous section it was pointed out that there are, in general, five tangents to a Δ^5 from any point of the plane. From a node, however, two pairs of tangents coincide and so there are only three tangents. Hence, if

$$t^5 + \alpha\mu t^4 + xt^3 + \alpha y t^2 + \mu t + \alpha = 0 \tag{1}$$

be considered a polynomial in t, with roots t_i (where $i = 1, 2, \cdots, 5$), then, at a node,

$$t_1 = t_2, \qquad t_3 = t_4.$$

Equating the symmetric functions of these roots to the coefficients of (1), we have

$$2\sigma_1 + t = -\alpha\mu, \qquad 2\sigma_1 t + \sigma_1^2 + 2\sigma_2 = x,$$

$$2\frac{\sigma_1}{\sigma_2} + \frac{1}{t} = -\frac{\mu}{\alpha}, \qquad \sigma_2^2 t = -\alpha,$$

where

$$t \equiv t_5, \qquad \sigma_1 \equiv t_1 + t_3, \qquad \sigma_2 \equiv t_1 t_3.$$

If now we put $t_1 \equiv 1/t$ and $\kappa \equiv 1/-\alpha$ and eliminate σ_1 and σ_2 from the third equation by means of the first and fourth equations, we obtain

$$\kappa t_1^5 - \kappa^3 \mu t_1^3 + \mu t_1^2 - \kappa^2 = 0. \tag{6}$$

Or by eliminating σ_1 and σ_2 from the second equation, with the help of the first and fourth, there results

$$4x = -3t_1^4 + 2\kappa^2 \mu t_1^2 + \kappa^4 \mu^2 + 8\frac{\kappa}{t_1}. \tag{7}$$

Of the three tangents from a node, only one is not tangent at the node, this we shall call the *odd* tangent. Then equation (6) gives the odd tangents. Evidently there are, in general, five such tangents. If $\mu > 10$, at least two of these tangents disappear.

Equation (7) represents a curve on which the five nodes lie. By means of relation (6) it can be brought to a simpler form, viz.

$$4x = -\mu\frac{\kappa^3}{t_1^3} + (\mu^2 + 5)\frac{\kappa}{t_1} + 3\mu\frac{t_1}{\kappa}.$$

Since it is always allowable to substitute one turn for another in the map equation of a curve, we can here put $\tau \equiv \kappa/t_1$ and thus obtain

$$4x = -\mu\tau^3 + (\mu^2 + 5)\tau + \frac{3\mu}{\tau}. \tag{8}$$

This is the equation of a parastroid,* or parallel to an astroid, which is not changed by varying $\alpha(=-\kappa^2)$; therefore, *the locus of nodes of the system of Δ^5's having an inscribed ellipse in common is a concentric parastroid.*

The line equation of this curve is

$$\tau^4 - \frac{2}{\mu}x\tau^3 + \left(\mu + \frac{5}{\mu}\right)\tau^2 - \frac{2}{\mu}y\tau + 1 = 0. \tag{9}$$

Since the coefficient of τ^2 is a constant times the distance between a tangent of the parastroid and a corresponding parallel tangent of its astroid, the shape of (9) depends wholly on the value of $\mu + 5/\mu$. The curve is an astroid when

$$\mu + \frac{5}{\mu} = 0, \qquad \text{or} \qquad \mu = \sqrt{-5},$$

but by hypothesis μ is real, therefore (9) can never be an astroid. The curve is tangent to itself, if

$$\mu + \frac{5}{\mu} = \pm 2,$$

but this equation again has no real roots. The real limitations of the curve (9), however, can best be seen from the equation giving the parameters of the cusps, which is obtained on equating to zero the derivative of x as to τ in equation (8). It is

$$3\mu\tau^4 - (\mu^2 + 5)\tau^2 + 3\mu = 0.$$

Two cusps coincide when this has equal roots, that is, when

$$(\mu^2 - 1)(\mu^2 - 25) = 0,$$

or

$$\mu = 1, \qquad \mu = 5.$$

So there are 5 cusps for $1 < \mu < 5$. For the limiting values, 1 and 5, the cusps coincide by twos and we get the form† of the parastroid which closely resembles an ellipse whose axes are to each other as one to two. For $\mu > 5$ and for $0 < \mu < 1$, the curve (9) has no cusps and varies in shape from the form just mentioned (where $\mu = 1$ or 5) to the circle (where $\mu = 0$ or ∞).

* G. LORIA: *Spezielle ebene Kurven*, p. 651.

† WOLSTENHOLME'S *Mathematical Problems*, p. 303.

The odd tangents from the nodes, we have seen, must satisfy relation (6) above. The clinant of any one of these lines is

$$-\frac{\alpha}{t}=\frac{\kappa^2}{t_1^2}=\tau^2,$$

these turns having the same value as they had above. A clinant of a normal to the node locus (8) at a node is

$$-\frac{dx}{dy}=\frac{\tau^2\left[3\mu\tau^4-(\mu^2+5)\tau^2+3\mu\right]}{\left[3\mu-(\mu^2+5)\tau^2+3\mu\tau^4\right]}=\tau^2.$$

Since the odd tangent and the normal to the node locus at a node have the same clinant and pass through the same point, they must coincide. So we may say that *the five odd tangents of any* Δ^5 *are normal to the locus of nodes of the system of* Δ^5*'s which have an inscribed ellipse in common.*

§ 5. *Cusps, cusp loci, and cuspidal tangents.*

The map equation of the Δ^5 is

$$x=-\left(3t^2+2a\mu t-\frac{\mu}{t^2}-\frac{2\alpha}{t^3}\right). \tag{2}$$

A cusp of this curve will be such a point, x, that for a change in t there will be no change in x; or, borrowing a phrase from kinematics, a cusp is that point of the curve at which the generating point has zero velocity. The condition, therefore, is $dx/dt=0$. If the values of t thus obtained are turns, then there are cusps. From (2) we get as the cusp equation

$$3t^5+\alpha\mu t^4+\mu t+3\alpha=0, \tag{10}$$

the roots of which are, in general, turns; hence we say the Δ^5 has *five cusps*. If, however, $\mu>15$, at least two of these cusps must disappear.

If equation (10) be subtracted from three times equation (1), we get

$$2\alpha\mu t^3+3xt^2+3\alpha yt+2\mu=0,$$

or

$$3x=-2\mu\left(2\alpha t-\frac{1}{t^2}\right),$$

the equation of a concentric deltoid; that is,* *the five cusp-tangents of a* Δ^5 *touch a concentric* Δ^3. And in general, by a similar argument, we see that *the* $2n-1$ *cusp-tangents of a* Δ^{2n-1} *touch a concentric* Δ^{2n-3}.

*The proofs of these three theorems are due to Professor MORLEY, v. Transactions of the American Mathematical Society, vol. 4 (1903), p. 8. The first two theorems are given by CLIFFORD, loc. cit.

If (10) be subtracted from (1), there results

$$2t^5 - xt^3 - \alpha y t^2 + 2\alpha = 0,$$

or

$$x = 2\left(3t^2 - \frac{\alpha}{t^3}\right),$$

the equation of a concentric regular Δ^5. *The five cusp-tangents of a* Δ^5 *touch a concentric regular* Δ^5. In a similar manner, *the* $2n-1$ *cusp-tangents of a* Δ^{2n-1}, *determined by* $2n$ *constants, touch a concentric* Δ^{2n-1} *determined by* $2n-2$ *constants.*

The cusp locus for the system of Δ^5's, having the same inscribed ellipse, is obtained by eliminating α between equations (10) and (2). It is

$$x = -6t^2 + \frac{(3t^4+\mu)(3\mu t^4+1)}{(\mu t^4+3)t^2},$$

or

$$x = -6\tau + \frac{(3\tau^2+\mu)(3\mu\tau^2+1)}{(\mu\tau^2+3)\tau}, \tag{11}$$

where $\tau = t^2$. This is a rational curve of the sixth order. So then, *the cusps of a system of* Δ^5*'s having a common inscribed ellipse lie on a rational curve of the sixth order.*

The cusp-tangent of any cusp of the system is obtained by eliminating α between equations (10) and (1). The result is

$$2\mu\tau^4 - (\mu x - 3y)\tau^3 + (\mu y - 3x)\tau - 2\mu = 0, \tag{12}$$

or

$$2\mu(\tau^4-1) - (\mu\tau^2+3)\tau x + (\mu+3\tau^2)\tau y = 0, \tag{13}$$

where $\tau \equiv t^2$. What this line envelopes as α varies may be seen by the introduction of Cartesian coördinates, but a simpler method, though it is somewhat artificial, is as follows: Take an ellipse in the form

$$x = \tau - \frac{\mu}{3\tau}. \tag{14}$$

The clinant of a normal to this ellipse at any point x is

$$-\frac{dx}{dy} = -\frac{\frac{dx}{d\tau}}{\frac{dy}{d\tau}} = \frac{3\tau^2+\mu}{\mu\tau^2+3},$$

hence the equation of this normal is

$$2\mu(\tau^4-1) - (\mu\tau^2+3)x\tau + (3\tau^2+\mu)y\tau = 0,$$

which is obviously the same as equation (13). Hence, *the cusp-tangents of a system of Δ^5's having the same orthoptic curve are normals of a concentric ellipse; and, therefore, their envelope is the evolute of this ellipse.*

This is the *second ellipse* which we have found common to the Δ^5's of the system.

§ 6. *Some special cases.*

I. *Curve for $\mu = 1$.*

The particular pentadeltoid for which $\mu = 1$ stands out as a curve of special interest on account of the behavior of its nodes and odd tangents. Here equation (6) becomes

$$\kappa t_1^5 - \kappa^3 t_1^3 + t_1^2 - \kappa^2 = 0, \tag{6a}$$

or

$$(t_1^2 - \kappa^2)(\kappa t_1^3 + 1) = 0,$$

whose roots are obviously $\pm\kappa$ and the three cube roots of $-1/\kappa$. But $t_1^2 \equiv t$ and $\kappa^2 \equiv -\alpha$, therefore the first two roots are simply equivalent to $t = -\alpha$ in the equation

$$t^5 + \alpha t^4 + xt^3 + \alpha y t^2 + t + \alpha = 0, \tag{1a}$$

whence we have as the equation of the two odd tangents

$$x - y = 0,$$

the axis of reals. In this case therefore *two of the odd tangents coincide and remain fixed as α varies.* In fact, by putting $\mu = 1$ in equation (4), we find that two of the nodes are the fixed points ± 2, whatever the value of α.

Putting the three other roots of (6a) into the equation of the curve on which the nodes lie, that is in

$$4x = -\frac{\kappa^3}{t_1^3} + 6\frac{\kappa}{t_1} + 3\frac{t_1}{\kappa},$$

we obtain

$$4x_1 = \tau^3 - 6\tau - \frac{3}{\tau},$$

$$4x_2 = \tau^3 - 6\omega\tau - \frac{3\omega^2}{\tau},$$

$$4x_3 = \tau^3 - 6\omega^2\tau - \frac{3\omega}{\tau},$$

where τ, $\omega\tau$, and $\omega^2\tau$ are the three cube roots of κ^4. Now the odd tangents from these nodes are readily obtained, since a point and a clinant of each is known. The clinants are

$$-\frac{\alpha}{t} = \frac{\kappa^2}{t_1^2} = \sqrt[3]{\kappa^4} = \tau^2, \omega^2\tau^2, \omega\tau^2.$$

Hence the tangents are:

$$x - \tau^2 y = \tau^3 - \frac{1}{\tau},$$

$$x - \omega^2 \tau^2 y = \tau^3 - \frac{\omega^2}{\tau},$$

$$x - \omega \tau^2 y = \tau^3 - \frac{\omega}{\tau}.$$

These three odd tangents evidently pass through the same point τ^3 and are equally spaced. This may be stated thus:

When $\mu = 1$, *three of the odd tangents pass through a point and are equally spaced.* (See Fig. 3.)

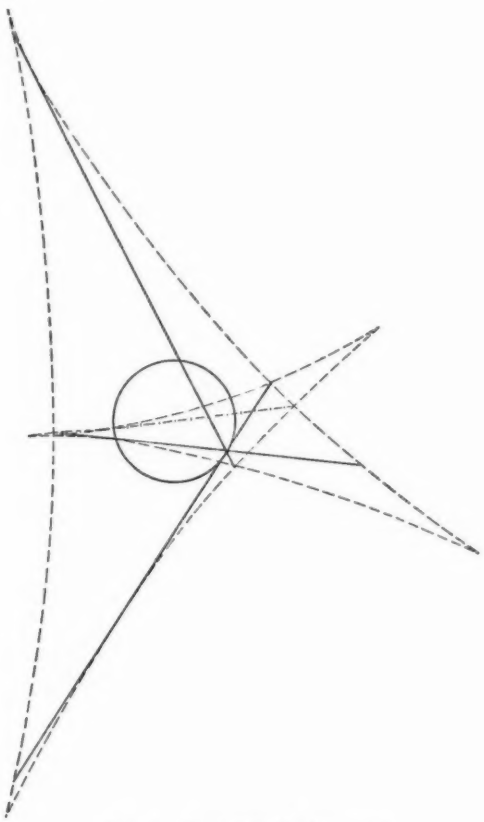


FIG. 3. Pentadeltoid when $\mu = 1$.

Since $\tau^3 = \kappa^4 = \alpha^2$, the point τ^3 is

$$x_1 = \alpha^2,$$

which for varying α traces out the unit circle. The point α^2 is an equiangular point of the triangle formed by the three nodes x_1, x_2, and x_3, and may be shown to be the *positive* one. It follows that *the triangle formed by the three variable nodes is such that as α varies, its vertices trace out the parastroid*

$$4x = \tau^3 - 6\tau - \frac{3}{\tau}, \tag{8a}$$

and its positive equiangular point moves on the unit circle.

From the point α^2, there will be two tangents to the Δ^5 other than the odd tangents mentioned above. The parameters of these are obtained by putting $x = \alpha^2$ in equation (1*a*). In the resulting equation

$$(t^2 + \alpha t + \alpha^2)\left(t^3 + \frac{1}{\alpha}\right) = 0,$$

the second factor gives the tangents already found. The roots given by the first factor are

$$t_1 = \omega\alpha \qquad \text{and} \qquad t_2 = \omega^2\alpha.$$

These values of t, substituted in (1*a*), give as the equation of the two tangents wanted:

$$x + \omega^2 y = \alpha^2 + \frac{\omega^2}{\alpha^2}, \qquad x + \omega y = \alpha^2 + \frac{\omega}{\alpha^2},$$

two lines whose directions are independent of the value of α, and which, together with the axis of imaginaries, form an equilateral triangle. *As α varies, these two tangents through the point α^2 but not through nodes move parallel to themselves and together with the axis of imaginaries form an equilateral triangle.*

II. *Curve for $\mu = 3$.*

This special case is of interest on account of its cusps and its cusp-tangents. The equations giving the parameters of the cusps can be solved when $\mu = 3$. For convenience, however, we shall take $\mu = -3$, that is, revolve our curve through an angle $\pi/2$. Then equation (10) becomes

$$3(t^5 - \alpha t^4 - t + \alpha) = 0, \tag{10b}$$

or

$$(t^4 - 1)(t - \alpha) = 0,$$

of which the roots are α, ± 1 and $\pm i$. The cusps are obtained by putting

these values of t in the map equation of the curve (2) when $\mu = -3$. Thus the five cusps are:

$$x_1 = 3\alpha^2 - \frac{1}{\alpha^2},$$

$$x_2 = 6 - 8\alpha, \qquad x_4 = -6 - 8i\alpha,$$

$$x_3 = 6 + 8\alpha, \qquad x_5 = -6 + 8i\alpha.$$

The cusp-tangent at the cusp x_1 is found by putting $t = -\alpha$ in the line equation of the Δ^5, that is, in

$$t^5 - 3\alpha t^4 + xt^3 + \alpha y t^2 - 3t + \alpha = 0. \tag{1b}$$

It is

$$x + y = 2\left(\alpha^2 + \frac{1}{\alpha^2}\right),$$

a line which is always parallel to the axis of imaginaries, since its clinant is independent of α. The cusp itself, as α varies, traces out an ellipse whose equation is

$$x = 3\alpha^2 - \frac{1}{\alpha^2},$$

which is recognized as the inscribed ellipse (3) when $\mu = -3$. Hence, *when $\mu = -3$, one cusp of the Δ^5 moves on the inscribed ellipse, while the cusp-tangent is always perpendicular to the axis of reals.*

The cusp-tangents at x_2 and x_3 are *

$$x + \alpha y = 2 + 2\alpha, \qquad x - \alpha y = 2 - 2\alpha,$$

two lines which are perpendicular to each other, and which always pass through the point 2, whatever the value of α. From the values of x_2 and x_3 given above it is evident that *the two cusps themselves move along a circle of radius 8 with center at the point 6, as α varies; in fact, they are opposite extremities of a diameter of this circle* (see Fig. 4).

In a similar manner it is found that *the other cusp-tangents are*

$$x - i\alpha y = -2 + 2i\alpha, \qquad x + i\alpha y = -2 - 2i\alpha,$$

two lines which pass through the point -2, and which are perpendicular to each other, for all values of α. These two cusps are opposite extremities of a diameter, tracing out the circle of radius 8 and center -6, (see Fig. 4).

When $\mu = -3$, therefore, the five cusps trace out the curves:

* Obtained by substituting $t = \pm 1$ in equation (1*b*).

$$x = 3\alpha^2 - \frac{1}{\alpha^2} \cdots \text{an ellipse};$$

$$x = 6 + 8\alpha \cdots \quad \text{a circle};$$

$$x = -6 + 8\alpha \cdots \quad \text{a circle}.$$

The ellipse (14), to which the cusp-tangents are normal, here degenerates into the segment

$$x = \tau + \frac{1}{\tau}, \tag{14b}$$

to which one cusp-tangent is always perpendicular, and through the extremities of which (the points ± 2), the four other tangents pass in pairs.

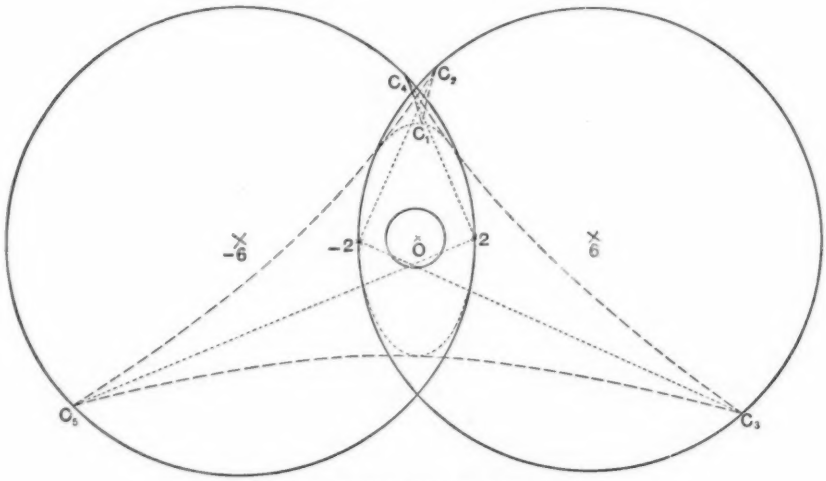


FIG. 4. Pentadeltoid when $\mu = -3$.

§ 7. *Three ellipses common to the system.*

Two ellipses which are common to all Δ^5's of the system (1), have already been found. The first is the *orthoptic ellipse*

$$x = -\left(t^2 + \frac{\mu}{t^2}\right), \tag{3}$$

or

$$x = \tau + \frac{\mu}{\tau};$$

the second is an ellipse to which all the cusp-tangents of the system are normal. The equation of this *second ellipse* is

$$x = \tau - \frac{\mu}{3\tau}. \tag{14}$$

A *third ellipse* arises in the following manner: A line cuts the Δ^5 in six points, whose parameters are found by substituting in the equation of a line,

$$\xi x + \eta y - 1 = 0,$$

values of x and y from the map equation of the Δ^5,

$$x = -\left(3t^2 + 2\alpha\mu t - \frac{\mu}{t^2} - \frac{2\alpha}{t^3}\right).$$

Thus a sextic in t is obtained, whose equation is

$$-2\frac{\eta}{\alpha}t^6 + (3\xi - \mu\eta)t^5 + 2\alpha\mu\xi t^4 + t^3 + 2\mu\frac{\eta}{\alpha}t^2 + (3\eta - \mu\xi)t - 2\alpha\xi = 0,$$

which may be represented symbolically by the form

$$(\alpha_1 t + \alpha_2)^6 \equiv (\beta_1 t + \beta_2)^6 = 0.$$

The six roots of this sextic will be a self-apolar set if

$$(\alpha_1\beta_2 - \alpha_2\beta_1)^6 = 0.$$

Putting the coefficients of the sextic into this apolarity condition, we obtain

$$4\xi\eta - \tfrac{1}{6}(3\xi - \mu\eta)(3\eta - \mu\xi) + \tfrac{4}{15}\mu^2\eta\xi - \tfrac{1}{40} = 0,$$

or

$$4(5\xi + \mu\eta)(5\eta + \mu\xi) = 1, \tag{15}$$

the line equation of a conic. Since ξ and η appear symmetrically, the curve is symmetrical with respect to the two axes and its center is the origin. Hence by making $\xi = \eta$, we find the points in which the conic cuts the real axis. They are $\pm(5 + \mu)$. Similarly by making $\xi = -\eta$, we have the intercepts on the axis of imaginaries. They are $\pm i(5 - \mu)$. These four points are actual intersections, therefore the conic is an ellipse whose map equation is easily seen (from its intersections on the axes) to be

$$x = 5\tau + \frac{\mu}{\tau}. \tag{16}$$

The geometrical significance of the parameters being self-apolar is that the tangents of the curve at the six points on a line cut the line at infinity in six

self-apolar points. Hence, *lines which cut the Δ^5 in six points, the tangents at which cut the line at infinity in a self-apolar set, envelope a concentric ellipse which is the same for all Δ^5's of the system having a common inscribed ellipse.*

In general *lines which cut a Δ^{2n-1} in points whose parameters are a self-apolar set envelope a concentric ellipse.* This ellipse is not the same, however, for each Δ^{2n-1} of the system having the same orthoptic curve if $n > 3$. If $\mu = 0$, then (16) becomes

$$x = 5\tau,$$

the equation of a circle through the cusps. If $\mu = 5$, equation (15) becomes the equation of the two points ± 10, through which we have two pencils of

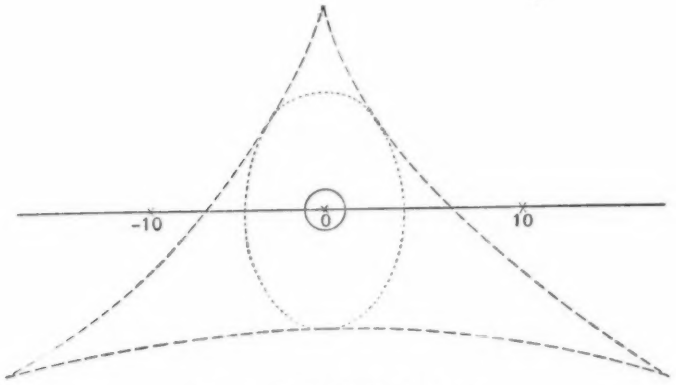


FIG. 5. Pentadeltoid when $\mu = 5$.

lines cutting the Δ^5 in six points whose parameters are a self-apolar set (see Fig. 5).

This *third ellipse of the Δ^5 is the same as the second ellipse of the evolute of the Δ^5.* This can be easily shown.

A normal to the Δ^5 is evidently

$$-xt^3 + \alpha yt^2 = t^3\left(3t^2 + 2\alpha\mu t - \frac{\mu}{t^2} - \frac{2\alpha}{t^3}\right) - \alpha t^2\left(\frac{3}{t^2} + \frac{2\mu}{\alpha t} - \mu t^2 - \frac{2t^3}{\alpha}\right),$$

whence

$$5t^5 + 3\alpha\mu t^4 + xt^3 - \alpha yt^2 - 3\mu t - 5\alpha = 0, \tag{17}$$

the equation of another Δ^5, whose orthoptic ellipse is

$$x = -\left(5t^2 - \frac{3\mu}{t^2}\right),$$

and whose second ellipse is

$$x = -\left(5t^2 + \frac{\mu}{t^2}\right), \qquad \text{or} \qquad x = 5\tau + \frac{\mu}{\tau}.$$

Comparing this with equation (16) we see it is the same as the ellipse which we have called the *third ellipse* of the original Δ^5; thus the theorem is proved.

Incidentally, we have proved above that *the evolute of a* Δ^5 *is another* Δ^5. So, in general, *the evolute of a* Δ^{2n-1} *is another* Δ^{2n-1}.

JOHNS HOPKINS UNIVERSITY,
March, 1905.

THE GROUPS OF ORDER p^m WHICH CONTAIN EXACTLY p CYCLIC SUBGROUPS OF ORDER p^α*

BY

G. A. MILLER

If a group (G) of order p^m contains only one subgroup of order p^α, $\alpha > 0$, it is known to be cyclic unless both $p = 2$ and $\alpha = 1$.† In this special case there are two possible groups whenever $m > 2$. The number of cyclic subgroups of order p^α in G is divisible by p whenever G is non-cyclic and $p > 2$.‡ In the present paper we shall consider the possible types of G when it is assumed that there are just p cyclic subgroups of order p^α in G. That is, we shall consider the totality of groups of order p^m which satisfy the condition that each group contains exactly p cyclic subgroups of order p^α. It is evident that $\alpha < m$.

Since the total number of subgroups of order p^α in G is of the form $1 + kp$, it follows that $\alpha > 1$. For all values of α greater than unity there is at least one group of order p^m which contains exactly p cyclic subgroups of order p^α, viz., the abelian group of type $(m-1, 1)$. When p is odd there is a non-abelian group which is conformal with this abelian group. It will be proved that these two groups are the only groups of order p^m, $p > 3$, which contain exactly p cyclic subgroups of order p^α. These two groups exist also when $p = 3$ or 2 and $m > 3$, but they are not the only groups which contain p cyclic subgpoups of order p^α, $p = 2$. When $\alpha = 2$ and $m = 4$ there is another group of order 3^m which contains just 3 cyclic subgroups of order 9.

Let the p cyclic subgroups of order p^α be represented by $P_1, P_2, \cdots, P_p$. Each of these transforms every other one into itself. The group generated by any two of them contains all the others, since it cannot be cyclic. Let s_1, s_2, be generators of P_1, P_2, respectively. From the fact that $s_1^{-1} s_2 s_1 = s_2^\beta$ and $s_2^{-1} s_1 s_2 = s_1^\gamma$, it follows that the commutator subgroup of the group (K) generated by s_1, s_2, is composed of operators which are invariant under this group. The order of this commutator subgroup cannot exceed p since $s_1^p = s_2^{p^p}$. This

* Presented to the Society (Chicago) December 30, 1905. Received for publication December 28, 1905.

† BURNSIDE, *Theory of groups of finite order*, 1897, p. 75.

‡ Proceedings of the London Mathematical Society, vol. 2 (1904), p. 142.

last equation results directly from the fact that G contains only p cyclic subgroup of p^α; for if it were not satisfied, s_1^p and s_2 would generate a group which would contain at least p cyclic subgroups of order p^α without containing s_1.

From the given equations it follows that the order of K is $p^{\alpha+1}$ and that K is one of the two groups of this order which contain p cyclic groups of order p^α. The only exception to this is when both $\alpha = 2$ and $p = 2$. In this special case K is completely determined by the given conditions, being the abelian group of order 8 and of type (2, 1). This establishes the following theorem: *If a group of order p^m contains exactly p cyclic subgroups of order p^α, these subgroups generate a characteristic subgroup of order $p^{\alpha+1}$, which is either the abelian group of type $(\alpha, 1)$, or the non-abelian group which is conformal with this abelian group.*

We shall now prove that K is abelian whenever G contains operators whose order exceeds p^α. If s is such an operator it may be assumed without loss of generality that $s^p = s_1$. Since there are only $p-1$ other cyclic subgroups of order p^α it follows that $s^{-1}s_2 s = cs_2 = s_2^\gamma$. The order of c cannot exceed p as $s_2^p = s_1^{\gamma p} = s^{\alpha p^2}$. Hence $s^p = s_1$ is commutative with s_2 or K is abelian. This theorem applies to every value of p. It should be observed that K contains just $p+1$ subgroups of every order. As each of the non-cyclic subgroups in K is characteristic, it follows that any operator (t) of order p^γ, which transforms K into itself, transforms any operator s of K such that if $t^{-1}st = c_1 s$, $t^{-1}c_1 t = c_2 c_1$, $t^{-1}c_2 t = c_3 c_2$, $\cdots$, then the order of c_1 is less than that of s, the order of c_2 is less than that of c_1, $\cdots$. When $c_{\alpha-1}$ is of order p^2 and c_α is of order p it is possible that $c_{\alpha+1}$ is also of order p. This special case will be considered in what follows.

Let t be any operator of order p^γ, $\gamma < \alpha$, which transforms K into itself and such that t^p is in K, and consider the order of the product ts_1, where s_1 has the same meaning as above. We have

$$(ts_1)^p = ts_1 ts_1 ts_1 \cdots p \text{ times} = ts_1 t^{-1} t^2 s_1 t^{-2} t^3 s_1 t^{-3} \cdots t^p$$

$$= c_1 s_1 c_2 c_1^2 s_1 c_3 c_2^3 c_1^3 s_1 \cdots = s_1^p k,$$

where k is the product of operators of lower order contained in K whenever $p > 3$. Hence the order of ts_1 is the same as that of s_1, viz., p^α whenever $p > 3$. This proves that G contains no operators whose orders divide p^α except those which are included in K, unless $p = 3$ or 2.

When $p = 3$ the above equations remain true whenever $\alpha > 2$. That is, if a group of order 3^m contains only 3 cyclic subgroups of order p^α, $\alpha > 2$, the group generated by these cyclic subgroups includes all the operators of the group whose orders divide p^α. We shall now consider the case when G contains operators whose orders exceed p^α. We shall again assume that $s^p = s_1$. The

group (K_1) generated by s and K is known to be conformal with the abelian group of type $(\alpha+1, 1)$.

If G should contain an operator (t) of order $p^{\alpha+1}$ which is not included in K_1 it could be assumed that $t^{-1}Kt = K$, and that $t^{p^\alpha} = s_1^{p^\alpha}$. Just as above it may be seen that $(st)^{p^\alpha} = s^{p^\alpha}t^{p^\alpha}$ into operators of lower order contained in K. As K is abelian it follows that $(st)^{p^\alpha} = s_1^{p^\alpha}s_1^{\alpha p^\alpha}$ into operators of lower order. By taking $a = -1$ it results that st is of a lower order than s. As this is impossible, K_1 includes all the operators of G whose orders divide $p^{\alpha+1}$. Hence we have the important result: *If a group of order p^m, $p > 3$, contains exactly p cyclic subgroups of order p^α it is either the abelian group of type $(m-1, 1)$, or the non-abelian group which is conformal with this abelian group.* When $\alpha > 2$, this theorem has also been proved for groups of order 3^m.

We shall now consider the groups of order 3^m which contain exactly 3 cyclic subgroups of order 9. If such a group contains also operators of order 27 it is conformal with the abelian group of type $(m-1, 1)$. This statement may be proved as follows. Let s be such an operator of order 27 and suppose that $s^3 = s_1$. The group of order 81 generated by s, s_2 is clearly conformal with the abelian group of type $(3, 1)$. If G contained an operator (t) of order 3 which is not found in this subgroup but transformed this subgroup into itself we should have $t^{-1}st = cs$, where c is of a lower order than s. Hence also $t^{-1}s^pt = (cs)^p = c^ps^p$. As t is commutative with c^p it follows that ts^p is of order p^2, contrary to the hypothesis that G contains only 3 cyclic subgroups of order 9. This proves the theorem. If a group of order 3^m contains only 3 cyclic subgroups of order 9 but contains also operators of order 27, it contains exactly four subgroups of order 3.

Suppose that G should contain an operator (t_1) of order 27 which is not contained in the group generated by s, s_2 but transforms this group into itself. It may be assumed that $t_1^3 = s^{-3}$. We have

$$(ts)^3 = tststs = tst^{-1}t^2st^{-2}t^3st^{-3}t^3 = c_1sc_2c_1^2sc_3c_2^3c_1^3s\cdot t^3 = s^3t^3k,$$

where k is of lower order than s^3t^3 and is commutative with s^3 and t^3. Hence $(ts)^9 = 1$, as s^3 and t^3 are also commutative. From this and the preceding paragraph it follows that all the operators of order 27 which are found in G are included in the group generated by s, s_2. That is, if G contains only 3 cyclic subgroups of order 9 but contains also operators of order 27, it also contains just 3 cyclic subgroups of order 27, and hence just 3 cyclic subgroups of every order which divides p^{m-1} and exceeds p.

The two preceding paragraphs prove that if a group of order 3^m contains just 3 cyclic subgroups of order 9 it is either conformal with the abelian group of type $(m-1, 1)$ or it contains only operators of order 3 in addition to the 18 of order 9 which are found in the 3 cyclic subgroups of order 9. In the latter

case its order is 81: for if its order exceeded 81 each cyclic subgroup of order 9 would be transformed into itself by at least 81 operators of G. There would therefore be operators of order 3 which would transform each cyclic subgroup of order 9 into itself but would not be in the group of order 27 generated by these cyclic subgroups. As such an operator would give rise to additional operators of order 9 this is impossible. Hence we have the result: If a group of order 3^m contains exactly 3 cyclic subgroups of order 3^α, $\alpha > 2$, it contains exactly 3 cyclic subgroups of each of the orders $9, 81, \cdots, 3^{m-1}$, and hence is conformal with the abelian group of type $(m-1, 1)$. When $\alpha = 2$ and $m > 4$ the same conclusions hold. When $\alpha = 2$ and $m = 4$ there is another group which contains exactly 18 operators of order 9, viz., the group which contains only operators of order 3 besides the identity and these operators of order 9.

Combining these results with those which precede we have that a group of order p^m, $p > 2$, which contains just p cyclic subgroups of order p^α contains just p cyclic subgroups of each of the orders $p^2, p^3, \cdots, p^{m-1}$. The only exception to this which may arise is when $p = 3$ and $m = 4$. In this special case there is a group which contains just p cyclic subgroups of order p^2 without also containing any cyclic subgroups of order p^3. In this case, there are therefore three groups of order p^m which contain just p cyclic subgroups of order p^α while in all other cases there are only two such groups. It remains to consider the cases when $p = 2$.

We shall first prove that if a group of order 2^m contains just two cyclic subgroups of order 2^α, $\alpha > 2$, it cannot contain more than two cyclic subgroups of any higher order. It has already been proved that these two cyclic subgroups generate a group K of order $2^{\alpha+1}$ and that K is abelian whenever G contains operators of order $2^{\alpha+1}$. Suppose that t_1 is an operator such that $t_1^2 = s_1$. The group generated by t_1, s_2 is of order $2^{\alpha+2}$ and is either abelian or contains a commutator subgroup of order 2, generated by s_1.

If t_2 is another operator of order $2^{\alpha+1}$ contained in G we may assume that it transforms the given subgroup of order $2^{\alpha+2}$ into itself since G contains at least one operator of order $2^{\alpha+1}$ which has this property as every subgroup is invariant under a larger subgroup. We may assume that $t_2^4 = s_1^{-2}$. Hence $(t_1 t_2)^2 = t_1 t_2 t_1 t_2^{-1} t_2^2 = c t_1^2 t_2^2$, where c, t_1^2, t_2^2 are commutative and c is of a lower order than t_1^2. The order of $t_1 t_2$ is therefore less than 2^α. As $t_1 t_2$ transforms an operator of order 2^α in K into itself multiplied by an operator whose order does not exceed 2, the group generated by K and $t_1 t_2$ would contain more than two cyclic subgroups of order 2^α. As this is impossible it has been proved that a group of order 2^m which contains only two cyclic subgroups of order 2^α, $\alpha > 2$, contains at most two cyclic subgroups of order $2^{\alpha+1}$. If it contains only two such subgroups they generate a group of order $2^{\alpha+2}$ which is either abelian or contains a commutator subgroup of order 2.

We proceed to prove the theorem: *If a group of order 2^m contains exactly two cyclic subgroups of order 2^β but no cyclic subgroup of any higher order, then $m \leqq 2^{\beta+2}$.* It has been proved that the two cyclic subgroups of order 2^β generate a group of order $2^{\beta+1}$ whose commutator subgroup is generated by $s_1^{2^{\beta-1}}$. Suppose that $m > 2^{\beta+2}$ and let t_1, t_2 be any operators of G which are not also in K, $t_1^{-1} s_1 t_1 = c_1 s_1$, $t_2^{-1} s_1 t_2 = c_2 s_1$. The orders of c_1, c_2 cannot be less than $2^{\beta-1}$, since G does not involve any operator of order 2^β besides those in K.* It has also been observed that these orders cannot exceed $2^{\beta-1}$ since t_1, t_2 transform K into itself. From $(t_1 t_2)^{-1} s_1 t_1 t_2 = c_3 s_1$, where c_3 is of a lower order than c_1, c_2, it follows that $t_1 t_2$ is in K. That is, the value of m does not exceed $2^{\beta+2}$.

The preceding results prove that if a group of order 2^m contains exactly two cyclic subgroups of order 2^α, $\alpha > 2$, it contains operators of order 2^{m-2} and hence has been determined.† It remains to consider the case when a group contains only two cyclic subgroups of order 4 and to prove that in this case it must also contain operators of order 2^{m-2}. If s_1 is an operator of largest order in such a G, the cyclic group which it generates is transformed into itself by each of the operators of order 4. Hence the group K generated by s_1 and these operators of order four is of order $2^{\beta+1}$, 2^β being the order s_1, and K is conformal with the abelian group of type $(\beta, 1)$.

If the order of G should exceed $2^{\beta+2}$, K would be transformed into itself by an operator t of order 8 such that the order of c in $t^{-1} s_1 t$ could not exceed $2^{\beta-2}$. The order of the product of t into some operator of order 8 in K could therefore not exceed 4. As this is impossible it has been proved that *every group of order 2^m which contains exactly two cyclic subgroups of a given order contains a cyclic subgroup of order 2^{m-2}.*

If we combine this result with those which precede we arrive at the theorem: Every group of order p^m, p being any prime, which contains exactly p cyclic subgroups of the same order must contain a cyclic subgroup of order p^{m-2}. When p is odd and $m > 4$, we have the stronger theorem: Every group of order p^m which contains exactly p cyclic subgroups of the same order contains exactly p cyclic subgroups of every order from p^2 to p^{m-1}. This theorem is also true when $m = 3$, and when $m = 4$ and $p > 3$. As all the groups of order p^m which contain a cyclic subgroup of order p^{m-2} are known, these results give a complete determination of all the groups of order p^m which contain exactly p cyclic subgroups of the same order.

*Cf. Bulletin of the American Mathematical Society, vol. 7 (1901), p. 351.

†Transactions of the American Mathematical Society, vol. 2 (1901), p. 259; Bulletin of the American Mathematical Society, vol. 9 (1905), p. 494.

GROUPS IN WHICH A LARGE NUMBER OF OPERATORS MAY CORRESPOND TO THEIR INVERSES*

BY

W. A. MANNING

An abelian group may be defined by the property that, in an automorphism of the group, more than three fourths its operators may be placed in a one to one correspondence with their inverses.† It may be of interest to know the groups possessing the property that five eighths or more of the operators may be made to correspond to their inverses. The principal object of this paper, however, is to establish the following elementary theorem (I) and to illustrate the use that may be made of it in certain problems.

THEOREM I. *A group that has two invariant subgroups with nothing in common but the identity can be set up as a multiple isomorphism between two groups of lower order.*

Let a group (G) of order $k_1 k_2 x$ have the two invariant subgroups K_1 and K_2 of order k_1 and k_2 respectively. If K_1 and K_2 have only the identity in common, every operator of K_1 is commutative with every operator of K_2. It may be assumed that G is not merely the direct product of K_1 and K_2. Let $1, r_2, r_3, \cdots$, be the operators of K_1 and $1, s_2, s_3, \cdots$, those of K_2. Now

$$
\begin{array}{ll}
(1_a) & (1_b) \\
G = K_2 + r_2 K_2 + \cdots + r_{k_1} K_2 & G/K_2 = K_1 \\
+ t_2 K_2 + t_2 r_2 K_2 + \cdots + t_2 r_{k_1} K_2 & + t_2' K_1 \\
+ t_3 K_2 + t_3 r_2 K_2 + \cdots + t_3 r_{k_1} K_2 & + t_3' K_1 \\
+ \quad \cdots \quad & + \cdots \\
+ t_x K_2 + t_x r_2 K_2 + \cdots + t_x r_{k_1} K_2, & + t_x' K_1,
\end{array}
$$

* Presented to the Society (San Francisco) September 30, 1905. Received for publication November 19, 1905.

† MILLER, Annals of Mathematics, ser. 2, vol. 7 (1906), p. 55.

$$
\begin{array}{ll}
(2_a) & (2_b) \\
G = K_1 + s_2 K_1 + \cdots + s_{k_2} K_1 & G/K_1 = K_2 \\
+ t_2 K_1 + t_2 s_2 K_1 + \cdots + t_2 s_{k_2} K_1 & + t_2'' K_2 \\
+ t_3 K_1 + t_3 s_2 K_1 + \cdots + t_3 s_{k_2} K_1 & + t_3'' K_2 \\
+ \cdots & + \cdots \\
+ t_x K_1 + t_x s_2 K_1 + \cdots + t_x s_{k_2} K_1, & + t_x'' K_2.
\end{array}
$$

The operators t_i' and t_i'' $(i = 2, 3, \cdots, x)$ correspond to $t_i K_2$ and $t_i K_1$ in (1_a) and (1_b) respectively. Let $\{K_1, K_2\}$ be called H. It is clear that $(G/K_2)/K_1 = (G/K_1)/K_2 = G/H$. We may write G/K_2 and G/K_1 as substitution groups on different sets of letters and by multiplying corresponding divisions in the order given in (1_b) and (2_b) obtain

$$G' = H + t_2' t_2'' H + t_3' t_3'' H + \cdots + t_x' t_x'' H.$$

That G' is simply isomorphic to G remains to be proved. The order of G' is clearly $k_1 k_2 x$. If in G, $t_i r_\alpha s_\lambda t_j r_\beta s_\mu = t_k r_\gamma s_\nu$, from (1_a) and (1_b) $t_i' r_\alpha t_j' r_\beta = t_k' r_\gamma$ and from (2_a) and (2_b) $t_i'' s_\lambda t_j'' s_\mu = t_k'' s_\nu$. Hence in G', $(t_i' t_i'' r_\alpha s_\lambda)(t_j' t_j'' r_\beta s_\mu) = t_i' r_\alpha t_j' r_\beta t_i'' s_\lambda t_j'' s_\mu = t_k' r_\gamma t_k'' s_\nu = t_k' t_k'' r_\gamma s_\nu$ and the truth of our theorem is now evident.

COROLLARY I. If an invariant subgroup of G has nothing but the identity in common with the commutator subgroup of G, G is a multiple isomorphism between an abelian and a non-abelian group.

COROLLARY II. Groups whose invariant operators are not all powers of a single one among them can be set up by establishing isomorphisms between groups of lower order.

COROLLARY III. All metabelian* groups of order p^m can be constructed by setting up isomorphisms between metabelian groups of lower order in which the groups of invariant operators are cyclic.

THEOREM II. *Given a group G in which a certain fractional part of the operators may be made to correspond to their inverses, we can take any invariant subgroup H for which G/H is abelian, set up an isomorphism with any abelian group having the same quotient group, and, by multiplying corresponding divisions, construct a new group in which the same proportion of the operators may correspond to their inverses.*

Let G have an invariant subgroup H of order h and index x, and let G/H be abelian. Also let A be an abelian group with a subgroup B of order h and index x, such that A/B is simply isomorphic to G/H. We may write G and

* FITE, Transactions of the American Mathematical Society, vol. 3 (1902), pp. 331-353.

A as regular groups on different sets of letters, and then set up an intransitive group (G') by taking the direct product of H and B, and of the operators in corresponding divisions of the tails of G and of A. Now if t be an operator which transforms the given fractional part of the operators of G into their inverses, and if t_1 transforms all the operators of A into their inverses, the operators of G' can be arranged in $hx = g$ sets of b each, with one operator of G in each set, so that tt_1 transforms every operator of a set into its inverse whenever the accompanying operator from G goes into its inverse.

Let us assume that exactly $\frac{3}{4}g$ operators of G can be made to correspond to their inverses. Since it may be assumed that G is not abelian a substitution t can be selected from G that corresponds to t^{-1} and is not invariant in G. Let $t_i (i = 1, 2, \cdots, \frac{3}{4}g)$ be the totality of operators in G that correspond to their inverses. Form the $\frac{3}{4}g$ products tt_i. At least $\frac{1}{2}g$ of the products tt_i correspond to $(tt_i)^{-1}$, giving $t^{-1}t_i^{-1} = t_i^{-1}t^{-1}$, so that t is commutative with $\frac{1}{2}g$ operators of G, all of which correspond to their inverses. This subgroup (H) in which t is invariant is abelian, since all its operators correspond to their inverses. Next let s be an operator not in H and which corresponds to s^{-1}. Since half the products st_i correspond to their inverses, s is commutative with half the operators of H. Hence H includes a subgroup (F) of order $\frac{1}{4}g$ all of whose operators are invariant in G. The group of cogredient isomorphisms, G/F, is axial. The converse is true. We have

$$G = F + tF + sF + stF.$$

From this we get $s^{-1}ts = tc$, where c, the commutator, is an operator of F: $s^{-2}ts^2 = tc^2 = t$, so that $c^2 = 1$. Hence $t^{-1}st = sc$, and the commutator subgroup of G is of order 2.

It may be assumed that G is not a direct product of, or multiple isomorphism between, an abelian and a non-abelian group. Let us suppose that the order of F is $2^{m-2}r$, where r is an odd number. Since G/F is axial, the result of taking odd powers of $tF(sF)$ is merely a permutation of the operators in each division. Hence $tF(sF)$ involves operators of order a power of 2. But F is the direct product of two groups, K of order 2^{m-2}, and K' of order r. Then G in turn is the direct product of $\{K, t, s\}$ and K', contrary to the assumption just made. If a and b are two operators of F, and if $t^2 = a$, $b^2 = a^{-1}$, it follows that $(tb)^2 = 1$, so that $t(s)$ may be assumed to have been so selected that $t^2(s^2)$ is not the square of an operator of F. Then t^2 and s^2 are independent generators of F, or the identity. By virtue of theorem I, corollary I, F has only one operator of order 2, the commutator, and in consequence F is cyclic.*

We may have $G = \{t, s\}$, where $t^{2^{m-1}} = s^{2^{m-1}} = 1$, $t^2 = s^2$, $s^{-1}ts = t^{1+2^{m-2}}$. There is one such group for all values of m greater than 3. When $m = 3$, G

* BURNSIDE, *Theory of Groups*, 1897, p. 75.

is the quaternion or octic group. If $t^{2^{m-1}} = s^2 = 1$, $sts = t^{1+2^{m-2}}$, st is of order 2^{m-1} since $ststst = t^2 t^{2^{m-2}}$, and we have the same series of groups as that just given. There remains the hypothesis that $t^2 = s^2 = 1$. F is generated by an operator r of order 2^{m-2}, and $trt = srs = r$, $sts = tr^{2^{m-3}}$. There is one such group whenever m is greater than 3. All other groups in which $\frac{3}{4}g$ operators may correspond to their inverses may be formed by means of theorem II.*

Let us suppose if possible that $\frac{2}{3}g + x$ $(0 < x < \frac{1}{12}g)$ operators of G correspond to their inverses. A non-invariant operator t which corresponds to t^{-1} is commutative with at least $\frac{1}{3}g + 2x$ operators of G. In this subgroup H (of order $h = \frac{1}{2}g$), $\frac{2}{3}h + 2x$ operators, perhaps more, correspond to their inverses.

Suppose that just $\frac{3}{4}h$ operators of H correspond to their inverses. These groups H are known. If F_1 is the subgroup of invariant operators of H, and if H_1 is an abelian subgroup of half the order of H, then

$$H = F_1 + t_1 F + s_1 F_1 + s_1 t_1 F_1 \qquad \text{and} \qquad G = H + sH,$$

where s is any operator of $G - H$ that corresponds to s^{-1}. It is clear that s is not commutative with all the operators of H, for in that event $\frac{3}{4}g$ operators of G would correspond to their inverses. Now s is commutative with just half the operators of G. Hence s is commutative with just half the operators of H, which form a subgroup (K) of H. Can all the operators of F_1 be invariant in G? To answer this question we note that if one operator in a division of G with respect to F_1 corresponds to its inverse, all the operators of that division do so, since the operators of F_1 all do so. But no integral multiple of $\frac{1}{8}$ is equal to $\frac{2}{3}$ or lies between $\frac{2}{3}$ and $\frac{3}{4}$. Now s is commutative with half the operators of H, and is not commutative with all the operators of H_1 (H_1 includes F_1) so that there is an operator s_1' in the tail $H - H_1$, commutative with s. The number of operators in the set $H_1 s_1'$ commutative with s is the same as the number of operators in H possessing this property. Hence s is commutative with just half the operators of H_1. In the same way s is seen to be commutative with just half the operators of F_1. These operators, invariant in G, form a subgroup F'. If one operator in a division of G with respect to F' corresponds to its inverse all the operators of that division do so. Then we have to see if it is possible for $\frac{11}{16}g$ operators of G to correspond to their inverses. Let a be an operator of F_1 which is not in F': $sa \neq as$. Of the 8 divisions of $G - H$ with respect to F', not both the sets in a division of $G - H$ with respect to F_1 can have all operators corresponding to their inverses. For example, if st_1 corresponds to $(st_1)^{-1}$, $st_1 a$ does not correspond to $(st_1 a)^{-1}$. Hence, instead of 11, at most 10 divisions of G taken with respect to F' contain operators which correspond to their inverses.

If H is abelian, all its operators correspond to their inverses since more than half do so, and s is commutative with half the operators of H, so that $\frac{3}{4}g$ oper-

ators of G correspond to their inverses, contrary to hypothesis. Now more than two thirds and less than three fourths of the operators of H correspond to their inverses when this condition is imposed on G, the same is true of a subgroup H' of H, and so on indefinitely, an absurdity when G is of finite order.

Except when $g/h = 2$ and $h/h' = 3$, for some pair of groups in the chain $G, H, H', \cdots$ above, the preceding reasoning shows that if the number of operators which correspond to their inverses is just $\frac{2}{3}g$, the subgroup H is of index 3. Neglecting for the moment the exceptional case, we distribute the operators of G in three sets H, uH, vH, where u is an operator that corresponds to u^{-1}. Since u cannot be commutative with all the operators of H without making G abelian, v can correspond to v^{-1}, so that one half the operators of each set uH and vH correspond to their inverses. Half the operators of H are invariant in the entire group. They form a subgroup F, with respect to which the quotient group is the non-abelian group of order 6. Then G has an abelian subgroup of order $\frac{1}{2}g$. We may write

$$G = (F + sF + s^2F) + tF + tsF + ts^2F.$$

We have $t^{-1}st = ssa$, where a is a certain operator of F; $t^{-2}st^2 = s(sa)^3 = s$, whence $(sa)^3 = 1$; $t^{-1}sat = (sa)^2$. Put $sa = c$. *Then* $G = \{F, c\} + t\{F, c\}$, *and is a* $(k, 3)$ *isomorphism between an abelian group of order* $2k$ *and the non-abelian group of order* 6.

Let t be commutative with all the operators of a subgroup H of order $\frac{1}{2}g$, and let just two thirds of the operators of H correspond to their inverses. We are at liberty to assume that H is one of the groups determined in the preceding paragraph and that every non-invariant operator which corresponds to its inverse is invariant in a subgroup of order $\frac{1}{2}g$. Consider an operator u of H that corresponds to u^{-1} and that is commutative with only $\frac{1}{6}g$ operators of H. Since $\frac{1}{2}g$ operators of G are commutative with u, two thirds of the operators of $G - H$ are commutative with u. Now if v' be one of these operators of $G - H$ commutative with u, we see that u is commutative with the same number of operators in $v'H$ as in H, a contradiction.

If more than $\frac{5}{8}g$ and less than $\frac{2}{3}g$ operators of G correspond to their inverses, t is invariant in a subgroup H of index less than 4.

Let H be of order $h = \frac{1}{3}g$. More than $\frac{3}{4}h$ operators of H correspond to their inverses, so that H is abelian, and all its operators correspond to their inverses. We write G in the three sets H, uH, vH. In $G - H$ there are more than $\frac{7}{24}g$ and less than $\frac{1}{3}g$ operators that correspond to their inverses. If u corresponds to u^{-1} and no operator of vH does so, between $\frac{7}{12}h$ and $\frac{2}{3}h$ operators of H are invariant in G, an impossibility. If v also corresponds to its inverse, more than $\frac{7}{24}h$ and less than $\frac{1}{3}h$ operators of H are invariant in G. Hence H cannot be of index 3.

Let H be of order $\frac{1}{2}g$. There is an operator u in $G - H$ which corresponds to u^{-1}. It is commutative with half the operators of H which form a subgroup K, invariant in G. Then H is not abelian. We now write

$$G = K + tK + uK + utK.$$

Since t and u are both commutative with every operator of K, $3x$ operators of K, tK, uK, correspond to their inverses whenever x operators of K do so. The maximum number of operators of utK that correspond to their inverses is $\frac{1}{4}g - x$, since ut does not correspond to $t^{-1}u^{-1}$. Then $3x + \frac{1}{4}g - x > \frac{5}{8}g$, $2x > \frac{3}{4}h$. But this is absurd because H is not abelian. This proves that *if more than $\frac{5}{8}g$ operators of G correspond to their inverses, just $\frac{2}{3}g$, $\frac{3}{4}g$, or all do so.*

From the preceding paragraphs it is clear that if just $\frac{5}{8}g$ operators of G correspond to their inverses, t cannot be commutative with just $\frac{1}{8}g$ operators of G. We first assume that there is an operator t which is invariant in a subgroup H of index 4, and not invariant in a larger subgroup. Then H is abelian. Write G in the sets H, uH, vH, wH. Just half the operators of a set uH correspond to their inverses. Since no operator from $G - H$ is commutative with t, just half the operators of H are invariant in G. Let F be the group composed of the invariant operators of G. When G is not the direct product an abelian and a non-abelian group its order is a power of 2. The operator of G/F that corresponds to tF is of order 2. The group of cogredient isomorphisms G/F is neither quaternion, cyclic, of type $(2, 1)$,* nor, when G/F is octic, is the operator corresponding to tF invariant. Let G/F be octic and arrange G thus:

$$G = F + s^2F + sF + s^3F + tF + ts^2F + tsF + ts^3F.$$

By hypothesis, $ts^2 \neq s^2t$, so that $t^{-1}st = s^3a$ (a belonging to F),

$$t^{-1}s^4t = s^{12}a^4 = s^4, \qquad (s^2a)^4 = 1, \qquad t^{-1}s^2at = (s^2a)^{-1}.$$

The commutator subgroup of G is cyclic of order 4. Let us suppose that the invariant abelian subgroup F has an operator b of order 2 other than $(s^2a)^2$, the commutator of order 2. Then $\{b\}$ and $\{s^2a\}$ are two invariant subgroups of G with nothing but the identity in common, and according to theorem I, corollary I, G is an isomorphism between two groups of lower order, one abelian and one non-abelian. Making the same assumption in regard to the non-abelian constituent we continue until we have a group G with only the one operator of order 2, $(s^2a)^2$, in F. But an abelian group of order 2^m with only one operator of order 2 is cyclic. Now let s^4 be taken as an operator of highest order in F. Since $\{F, s\}$ is cyclic, $t^{-1}st = s^3s^{4r}$, $t^{-1}s^4t = s^{12}s^{16r} = s^4$, $s^8 = 1$, and

* MILLER, Comptes Rendus de l'Académie des Sciences, vol. 128 (1899), p. 229.

$x=0$ or 1. In both cases the groups are known and five eighths of their operators can be made to correspond to their inverses. There remains the hypothesis that $s^4=1$. F is generated by an operator r of order 2^{m-3}, commutative with both s and t; $\{F, s\}$ is of type $(m-3, 2)$, and either $t^2=1$ or $t^2=r$. Now $(st)^2=stst=sts t^{-1}t^2=ss^3r_1t^2=r_1t^2$, where r_1 is of order 4. If $t^2=r$, $(st)^2=rr_1$, so that when $m>5$ st is of the same order as t, and $\{t\}$ and $\{st\}$ have half their operators in common. This is the G_3 given by Professor MILLER* in a memoir "On the Groups of order p^m which contain operators of order p^{m-2}". When $m=5$, $r^4=1$, $s^{-1}ts=ts^2r$, $(st)^2=1$, and this G_{32} has only two cyclic subgroups of order 8. We may suppose t replaced by $t'=st$. This replacement transforms $\{r, s\}$ in the same way that t does. If $t^2=r^{2^{m-3}}=s^4=1$, consider the subgroup G' of order 32 generated by $r_1=r^{2^{m-5}}$, s and t. It is the group last mentioned. One and only one group G of order 2^m can be constructed from G' by adjoining an operator r commutative with every operator of G' and such that its 2^{m-5}th power is r_1.

Let G/F be of type $(1, 1, 1)$. G cannot have just one commutator, for then every operator would be commutative with just half the operators of G, contrary to the assumption in regard to t. The commutators are invariant, since G is metabelian, and moreover are all of order 2. The commutator subgroup is then of order 4 or 8. Let it be of order 4. We may suppose G deprived of its abelian constituents so that F is the direct product of two cyclic groups of orders 2^α and 2^β. Then G is a $(2^\alpha, 2^\beta)$ isomorphism between two non-abelian groups G_1 and G_2 of orders $2^{\alpha+3}$ and $2^{\beta+3}$ respectively. Every operator of $G_1(G_2)$ is invariant in half of $G_1(G_2)$. But the group of cogredient isomorphisms of $G_1(G_2)$ cannot be of order 8,† from which we conclude that one fourth of the operators of $G_1(G_2)$ are invariant and three fourths correspond to their inverses. If we write G_1 and G_2 with respect to the subgroups composed of half the invariant operators and, setting up an isomorphism between the quotient groups of order 8, multiply corresponding divisions, three fourths of the operators of the resulting group can be made to correspond to their inverses unless invariant operators of one constituent be made to correspond to non-invariant operators of the other. There is one group of order 2^m, $m>3$, which we may use for G_1 and G_2 and if all possible values are given to α and β, there are $(m-3)/2$ or $(m-4)/2$ groups of order 2^m, as m is odd or even. If the commutator subgroup of G is of order 8, we have the same conditions until it comes to setting up the final isomorphism. Note that if between two of the constituent groups invariant operators of the one are multiplied only into invariant operators of the other,

* MILLER, Transactions of the American Mathematical Society, vol. 3 (1902), p. 383.

† YOUNG, American Journal of Mathematics, vol. 15 (1893), p. 71.

the resulting group has, not an axial commutator subgroup, but a commutator subgroup of order 2. Then we have to make invariant operators of one constituent correspond to non-invariant operators in both the other constituents. This can be done. The number of these groups of order 2^{m+3} is $(m^2-\alpha^2)/12$, when $m \equiv \pm\alpha \pmod 6$ and $0 \leqq \alpha \leqq 3$.

There remains the hypothesis that every non-invariant operator t that corresponds to its inverse is commutative with half the operators of G, which make up a subgroup H. The square of every operator that corresponds to its inverse is invariant. $G_1 = H + sH$ and if s corresponds to s^{-1}, s is commutative with half the operators of H which then form an invariant subgroup K. If x operators of K correspond to their inverses, we have $3x + (\frac{1}{4}g - x) \geqq \frac{3}{8}g$, so that $x = \frac{3}{16}g = \frac{3}{4}k$ and just one fourth of the operators of the division stK correspond to their inverses. The invariant operators of K are the invariant operators of G. The group of cogredient isomorphisms of G is abelian of type $(1, 1, 1, 1)$. We write $K = F_1 + t_1F_1 + s_1F_1 + s_1t_1F_1$, $G = K + tK + sK + stK$. To sts_1t_1 corresponds $t^{-1}s^{-1}t_1^{-1}s_1^{-1}$, or, since these 4 operators severally correspond to their inverses, $t^{-1}s^{-1}t_1^{-1}s_1^{-1} = s^{-1}t^{-1}s_1^{-1}t_1^{-1}$, whence, $tst^{-1}s^{-1} = t_1s_1t_1^{-1}s_1^{-1}$. Then the commutator subgroup of G is of order 2. Conversely when the commutator subgroup is of order 2 and t, s, t_1, s_1, severally correspond to t^{-1}, s^{-1}, t_1^{-1}, s_1^{-1}, sts_1t_1 corresponds to $(sts_1t_1)^{-1}$. These metabelian groups are known.*

* FITE, l. c.

STANFORD UNIVERSITY.

FINITE PROJECTIVE GEOMETRIES*

BY

OSWALD VEBLEN AND W. H. BUSSEY

By means of such a generalized conception of geometry as is inevitably suggested by the recent and wide-spread researches in the foundations of that science, there is given in § 1 a definition of a class of tactical configurations which includes many well known configurations as well as many new ones. In § 2 there is developed a method for the construction of these configurations which is proved to furnish all configurations that satisfy the definition. In §§ 4–8 the configurations are shown to have a geometrical theory identical in most of its general theorems with ordinary projective geometry and thus to afford a treatment of finite linear group theory analogous to the ordinary theory of collineations. In § 9 reference is made to other definitions of some of the configurations included in the class defined in § 1.

§ 1. *Synthetic definition.*

By a finite projective geometry is meant a set of elements which, for suggestiveness, are called points, subject to the following five conditions:

I. *The set contains a finite number* (> 2) *of points. It contains subsets called lines, each of which contains at least three points.*

II. *If A and B are distinct points, there is one and only one line that contains A and B.*

III. *If A, B, C are non-collinear points and if a line l contains a point D of the line AB and a point E of the line BC, but does not contain A, B, or C, then the line l contains a point F of the line CA* (Fig. 1).†

A plane ABC (A, B, C being non-collinear points) is defined as the set of all points collinear with a point A and any point of the line BC. It may be proved by III that a plane so defined has the usual projective properties. For example, a plane is uniquely determined by any three of its points which are non-collinear, and the line joining any two points of a plane is contained wholly in the plane.

A k-space is defined by the following inductive definition. *A point is a* 0*-space. If $A_1, A_2, \cdots, A_{k+1}$ are points not all in the same $(k-1)$-space, the*

* Presented to the Society April 22, 1905. Received for publication November 20, 1905.

† The figures used in this paper are not to be considered as having the properties which they have in ordinary euclidean geometry, unless such properties follow from the explicitly stated hypotheses.

set of all points collinear with the point A_{k+1} and any point of the $(k-1)$-space $(A_1, A_2, \cdots, A_k)$ is the k-space $(A_1, A_2, \cdots, A_{k+1})$. Thus a line is a

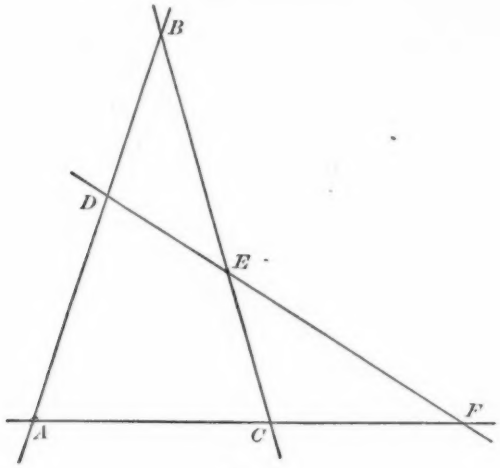


FIG. 1.

1-space, and a plane is a 2-space. From this definition it can be proved that spaces (if existent) satisfy the following well known theorem.

In a k-space, an l-space and an m-space have a point in common if $l+m \geqq k$. They have in common at least an r-space if $l+m-k=r$.

Remark: It is very convenient in practice to replace this theorem by a diagram consisting of $(k+1)$-points, of which any $l+1$ $(l \leq k)$ represent an l-space. Thus in a 4-space, using the diagram , it is evident that any two 3-spaces (each being a set of 4 points) have in common a plane (three points). This scheme gives a finite geometry satisfying all the projective geometry axioms except those implying that a line contains more than two points.

The existence of the various spaces is postulated by the two conditions IV_k and V_k. (Axioms of extension and closure.)

IV_k. *If l is an integer less than k, not all of the points considered are contained in the same l-space.*

V_k. *If IV_k is satisfied, there exists in the set of points considered no $(k+1)$-space.*

The well known *principle of duality* follows from the axioms and definitions as given. Special cases of the principle are the following:

Any proposition (deducible from $I-V_k$) about points and lines in a plane is valid if the words point and line be interchanged.

Any proposition (deducible from $I-V_k$) about points, lines, and planes in 3-space is valid if the words point and plane be interchanged.

Let $s+1$ denote the number of points in a line. To obtain the number of points in a plane consider a line l and a point L not in l. L and each point of l determine a line in each of which there are s points in addition to the point L. Furthermore, every point of the plane is in one of these $s+1$ lines containing L. Therefore the number of points in the plane is s^2+s+1. By the principle of duality the number of lines in the plane is also s^2+s+1. In like manner the number of points in 3-space may be found to be s^3+s^2+s+1 and the number of points in k-space to be $s^k+s^{k-1}+\cdots+s+1$.

The simplest example of a finite geometry which satisfies the definition for $k=2$ is the well known triple system

0	1	2	3	4	5	6
1	2	3	4	5	6	0
3	4	5	6	0	1	2

which consists of 7 points arranged in 7 lines of 3 points each. A finite plane geometry of 13 points arranged in 13 lines of 4 points each is the following:

0	1	2	3	4	5	6	7	8	9	10	11	12
1	2	3	4	5	6	7	8	9	10	11	12	0
3	4	5	6	7	8	9	10	11	12	0	1	2
9	10	11	12	0	1	2	3	4	5	6	7	8

For some purposes it is convenient to have the finite geometry exhibited in a table as follows:

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	x				x						x		x
1	x	x				x						x	
2		x	x				x						x
3	x		x	x				x					
4		x		x	x				x				
5			x		x	x				x			
6				x		x	x				x		
7					x		x	x				x	
8						x		x	x				x
9	x						x		x	x			
10		x						x		x	x		
11			x						x		x	x	
12				x						x		x	x

In this table the incidence of a line h and a point k is indicated by a mark x in the column h and the row k. The table shows at a glance the 4 points of a line and the 4 lines containing a point.

An example of a finite plane geometry having 5 points in a line may be found on p. 395 of vol. 5 of these Transactions.

350

§ 2. *Analytic definition.*

If $x_1, x_2, \cdots, x_{k+1}$ are marks of a Galois field* of order $s = p^n$, there are $(s^{k+1}-1)/(s-1) = s^k + s^{k-1} + \cdots + s + 1$ elements of the form $(x_1, x_2, \cdots, x_{k+1})$, provided that the elements $(x_1, x_2, \cdots, x_{k+1})$ and $(lx_1, lx_2, \cdots, lx_{k+1})$ are thought of as the same element when l is any mark $\neq 0$, and provided that the element $(0, 0, \cdots, 0)$ is excluded from consideration. These elements constitute a finite projective geometry of k-dimensions when arranged according to the following scheme. The equation

$$u_1x_1 + u_2x_2 + \cdots + u_{k+1}x_{k+1} = 0 \tag{1}$$

(the domain for coefficients and variables being the $GF[s]$) is said to be the equation of a $(k-1)$-space except when $u_1 = u_2 = \cdots = u_{k+1} = 0$. It is denoted by the symbol $(u_1, u_2, \cdots, u_{k+1})$. The symbols $(u_1, u_2, \cdots, u_{k+1})$ and $(lu_1, lu_2, \cdots, lu_{k+1})$, l being any mark $\neq 0$, denote the same $(k-1)$-space. The points of the $(k-1)$-space are those points of the finite geometry which satisfy its equation. A $(k-2)$-space is represented by two equations of type (1) and a $(k-l)$-space by l equations of type (1). There are $s^l + s^{l-1} + \cdots + s + 1$ points in l-space and, in particular, $s+1$ points in a line.

The finite projective k-dimensional geometry, obtained in this way from the $GF[s]$, is denoted by the symbol $PG(k, s)$. Since there is a Galois field of order s for every s of the form $s = p^n$, it follows that there is a $PG(k, p^n)$ for every pair of integers k and n and for every prime p. It will be proved in § 4 that every finite projective k-dimensional geometry satisfying the definition of § 1 is a $PG(k, p^n)$ if $k > 2$.

§ 3. *The modulus* 2.

The method used in § 2 to obtain the $PG(k, s)$ from the $GF[s]$ may be described as *analytic geometry in a finite field.* It may be applied to any field of finite order $s = p^n$, but here as elsewhere the modulus 2 gives rise to an exceptional case. Let the symbols 1, 3, 4, 5 denote the four vertices of a complete quadrangle (Fig. 2) in the $PG(2, p^n)$. Let the three pairs of opposite sides

* For the definition and properties of a Galois field see E. H. Moore, *Mathematical Papers, Chicago Congress of* 1893, pp. 210–226; L. E. Dickson, *Linear Groups*, pp. 1–14. For Galois field tables for $p^n \leqq 169$, $n > 1$ see W. H. Bussey, Bulletin of the American Mathematical Society, vol. 12 (1905–06), pp. 22–38.

meet as follows: 1 3 and 4 5 in a point 0, 1 4 and 3 5 in a point 2, and 1 5 and 3 4 in a point 6. Let the equations of the lines 1 5, 4 5, 3 4 be, respectively, $\alpha = 0$, $\beta = 0$, $\gamma = 0$ (abridged notation). As in SALMON'S *Conic Sections* (10th edition), p. 57, Ex. 1, the equations of the lines 0 6 and 2 6 may be found to be, respectively, $l\alpha + n\gamma = 0$ and $l\alpha - n\gamma = 0$, l and n being marks of the

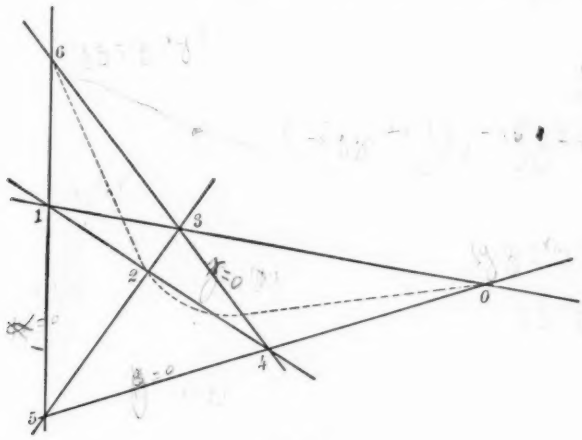


FIG. 2.

$GF[p^n]$. If the modulus of the field is 2, $n\gamma = -n\gamma$, and the lines 0 6 and 2 6 coincide, i. e., the diagonal points of the quadrangle are collinear. This does not happen if $p > 2$. It will be observed that the figure of the complete quadrangle in a geometry having the modulus 2 [i. e., in a $PG(k, 2^n)$] is thus proved to be identical with the triple system in 7 elements given in § 1.

To exclude from consideration the case of the modulus 2 it is therefore sufficient to add the following condition VI to those of the synthetic definition of § 1.

VI. *The diagonal points of a complete quadrangle are not collinear.*

In the following paragraphs, VI is not assumed unless it is so stated in the text.

§ 4. *General synthetic theory.*

The elementary part of the synthetic theory follows from conditions I–V quite independently of the hypothesis that the number of points is finite. If a certain further hypothesis is added, it is possible to develop* large parts of the theory

*The deduction from hypotheses I–VII of a portion of the usual projective geometry, corresponding in a general way to Part I of REYE'S *Geometrie der Lage* was carried out in detail by O. VEBLEN in a course of lectures delivered at the University of Chicago during the winter quarter, 1905. Mimeographed reports (referred to below as *Notes*) of these lectures as worked out by Mr. N. J. LENNES and other members of the class are on file in the mathematical library of the University of Chicago.

of collineations, conic sections, quadric surfaces, and algebraic curves and surfaces in general, without deciding whether the number of points is finite or not. This further hypothesis may be stated in the form:

VII. *Let A, B, C be three collinear points, and let A', B', C' be three other collinear points not on the same line. If the pairs of lines AB' and $A'B$, BC' and $B'C$, CA' and $C'A$ intersect, the three points of intersection are collinear.**

This hypothesis VII is a consequence of the hypothesis that the number of points is finite whenever the geometry considered is of three or more dimensions. The proof of this statement is in outline as follows: From hypotheses I–IV follows the DESARGUES theorem about perspective triangles: *If two triangles*

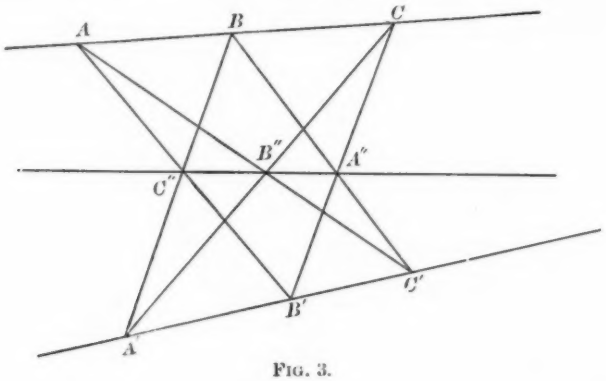


FIG. 3.

ABC and $A'B'C'$ are in the same plane and if the lines AA', BB', CC' are concurrent, the lines AB and $A'B'$, BC and $B'C'$, CA and $C'A'$ intersect in points which are collinear. This is proved on page 29 of the *Notes* referred to above, and by means of this theorem there is developed on pp. 73–84 a geometric algebra for the points of a line.† It is proved that this algebra satisfies all the conditions for a field except the commutative law of multiplication. In particular the algebra is such that for every element a there is a unique element a' such that $a'a = 1 = aa'$. For such an algebra it has been shown by J. H. MACLAGAN-WEDDERBURN‡ that the commutative law holds whenever

* The configuration involved here is known as the configuration of PAPPUS (Fig. 3). It may be described as a simple hexagon $AB'CA'BC'$ inscribed in two lines. Hypothesis VII is to the effect that the three pairs of opposite sides intersect, if at all, in three collinear points A'', B'', C''. HILBERT speaks of VII as *Pascal's theorem*. It is, of course, a degenerate case of the well-known theorem of PASCAL on a hexagon inscribed in a conic.

† Other developments of a geometric algebra practically equivalent to this one are given by G. HESSENBERG, Acta Mathematica, vol. 29 (1905), pp. 1–24, and by K. TH. VAHLEN, *Abstrakte Geometrie*, p. 110.

‡ *A theorem on finite algebras*, Transactions of the American Mathematical Society, vol. 6 (1905), p. 349.

the number of points is finite. On the other hand, it has been shown by HILBERT* (and it is shown by another method in the *Notes* referred to) that the commutative law of multiplication is equivalent to hypothesis VII. Therefore, if the number of points is finite, condition VII is satisfied.

By means of the algebra of points there may be built in the finite geometry defined by hypotheses I–VI a homogeneous analytic geometry of k dimensions in every case in which $k > 2$ (see pp. 82–84 of the *Notes*), and, since the algebra of points is abstractly identical with a Galois field, the analytic geometry so obtained is identical with that described in § 2. This proves that for a given $k > 2$, p and n there is one and only one finite projective geometry as defined in § 1 and that it is the $PG(k, p^n)$. (Cf. end of § 2.)

The essential difference in the case $k = 2$, i. e., in the case of plane geometry, is due to the fact that there exist finite non-desarguesian plane geometries.† The Desargues theorem about perspective triangles in a plane implies that the plane may be thought of as immersed in a 3-space. From this it follows that for a given p^n there is one and only one finite plane desarguesian geometry, viz., the $PG(2, p^n)$.

§ 5. *The Möbius net.*

On the line $x_1 = 0$, any point except (010) may be represented by a single coördinate, the value of x_2 when $x_3 = 1$. The point (010) may be represented

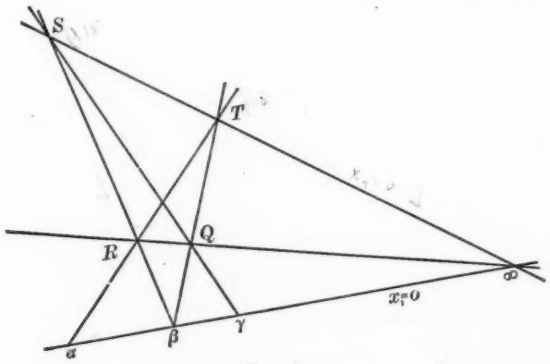


FIG. 4.

by the coördinate ∞. Let α and β be the coördinates of two points of the line. The harmonic conjugate of α with respect to β and ∞ may be constructed by the usual quadrangle construction as follows. Let S and T be any two points on a line through the point ∞. Let the lines $S\beta$ and $T\alpha$ meet in

* D. HILBERT, *Grundlagen der Geometrie*, chap. 6.

† Cf. *Non-desarguesian and non-pascalian geometries*, by J. H. MACLAGAN-WEDDERBURN and O. VEBLEN. This paper was read at the April (1905) meeting of the Chicago section but has not yet been published.

a point R. Let the lines $T\beta$ and $R\infty$ meet in a point Q. The line SQ meets the line $x_1 = 0$ in the required point γ (Fig. 4). If the line $ST\infty$ is chosen as the line $x_3 = 0$ and if S and T be chosen as the points $(\lambda 10)$ and (100) respectively, the equation of the line SQ may be proved to be $x_1 - \lambda x_2 + \lambda(2\beta - \alpha)x_3 = 0$. The point of intersection of the line SQ and the line $x_1 = 0$ is therefore $\gamma = 2\beta - \alpha$.

The well-known Möbius net * is determined by any three points of a line. If the three points be chosen as the points A_1, A_2, ∞ of the line $x_1 = 0$ (Fig. 5), the fourth point A_3 is determined as the harmonic conjugate of A_1 with respect to A_2 and ∞; A_4 is the harmonic conjugate of A_2 with respect to A_3 and ∞;

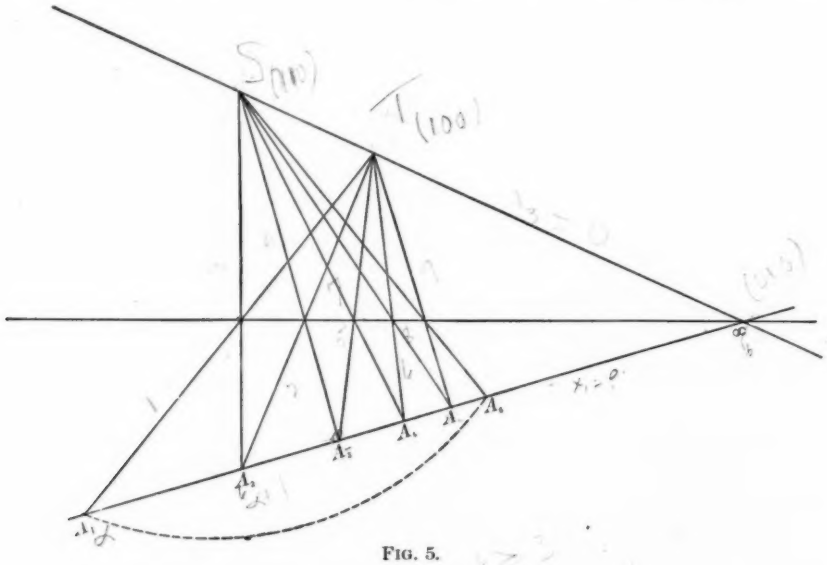


FIG. 5.

A_k is the harmonic conjugate of A_{k-2} with respect to A_{k-1} and ∞. The word net is used to denote the set of points A_1, A_2, $A_3 \cdots \infty$ and not to denote the whole figure.

Let the coördinate of the point A_1 be α and that of A_2 be $\alpha + 1$. Since the harmonic conjugate of a point α with respect to the points β and ∞ is the point $2\beta - \alpha$, the points of the net are α, $\alpha + 1$, $\alpha + 2$, $\cdots$, ∞. Since the modulus of the field is p, the series α, $\alpha + 1$, $\alpha + 2$, $\cdots$ consists of the p marks α, $\alpha + 1$, $\alpha + 2$, $\cdots$, $\alpha + p - 1$. The $(p + 1)$-st mark is again α, i. e., the series repeats itself periodically. Fig. 5 is drawn for the case $p = 5$. The

* KLEIN, *Nicht-Euclidishe Geometrie*, vol. 1, p. 338 ff. MÖBIUS, *Gesammelte Werke*, vol. 1, p. 237 ff.

point A_6 is to be thought of as coincident with the point A_1. This proves that a finite projective geometry cannot be represented by a figure in ordinary geometry in which a line of the finite geometry consists of a finite set of points on a line of ordinary geometry.

As in ordinary projective geometry it can be proved that any set of three collinear points may be projected into any other set of three collinear points. Therefore any two Möbius nets are projective and the number of points in any net in any line is $p+1$. A net is determined by any three of its points. The p^n+1 points of a line may be arranged in p^{n-1} nets of $p+1$ points each, all of the nets having a point in common.

A collineation is defined as a point transformation by which lines are transformed into lines. A collineation transforms four harmonic points into four harmonic points because it transforms a complete quadrangle into a complete quadrangle. Therefore, if a collineation leaves three points of a net fixed, it leaves the whole net fixed point by point. In the $PG(k, p)$ there are but $p+1$ points in a line, i. e., the points of a line constitute a single net. Therefore a collineation which leaves three points of a line fixed leaves the whole line fixed point by point. Therefore any collineation, other than the identity, has at most two fixed points in a line; in a plane it has at most three non-collinear fixed points; and in a k-space it has at most $k+1$ fixed points if no k of them are in the same $(k-1)$-space. In the $PG(k, p^2)$ a collineation may have more than three fixed points in a line, because when three points of the line are fixed only one net of the line is necessarily fixed point by point. The number of fixed points in this case is discussed below. The situation here is analogous to that in ordinary complex projective geometry. The nets correspond in their definition and properties to VON STAUDT'S* chains, i. e., to circles if the chains be represented in the complex plane. Any three points of a line in the $PG(k, p^2)$ determine a net or, as it is now to be called, a chain. Let α be a chain, and let P be a point of the line but not of the chain α. Let A be a point of the chain α. If X be any other point of α, the three points P, X, A determine one and only one chain. Since X may be chosen as any one of the $p+1$ points of α other than the point A, there are at least p chains containing P and A. The total number of points in these p chains is equal to $p(p-1)+2=p^2-p+2$. The number of points of the line which are not contained in any of these chains is therefore $(p^2+1)-(p^2-p+2)=p-1$. Let β be the chain determined by P, A, and any one of these $p-1$ points. Since β cannot contain any other point of any of the p chains through P and A, it must contain all of the above mentioned $p-1$ points. The chains α and β, therefore, have one and only one point in common and are said to be tangent to each other. We have thus proved that through a given point P of the line

* V. STAUDT, *Beiträge zur Geometrie der Lage*, p. 137.

there is one and only one chain tangent to a given chain α. The system of all chains having in common a given point of the line affords, if the given point be excluded, a system of sets of p points each such that no two sets have more than one point in common and such that any two points of the line determine one of the sets. Moreover, containing a given point of the line, there is one and only one set β which has no point in common with a given set α. Thus there is defined a set of points and subsets of it entirely analogous in their intersectional properties to the lines of the ordinary euclidean plane. It is also analogous to the system of circles through a fixed point of the ordinary complex plane. The system of all chains of a line is of course analogous to the system of all circles and straight lines in a plane.*

If a collineation leaves fixed, point by point, a chain α and a point P not in α, it must be the identity because through every other point of the line there passes at least one chain which contains P and two points of α, i. e., every other point of the line is contained in at least one chain of which three points are fixed. Therefore a collineation in the $PG(k, p^2)$ may have more than two fixed points on a line but not so many as four unless they are contained in the same chain.

In the $PG(k, p^n)$, the chains of a line constitute a configuration analogous to the set of all circles in n-dimensional space. Any four points of the line, if they are not all contained in the same chain, determine a 2-chain, a configuration analogous to a sphere. Any five points, if they are not all contained in the same 2-chain, determine a 3-chain, a configuration analogous to a hypersphere;† and any $(l+1)$ points $(l \leqq n)$, if they are not contained in the same $(l-2)$-chain, determine an $(l-1)$-chain, a configuration analogous to an $(l-1)$-sphere. A collineation can have as fixed points at most $n+1$ points not all contained in the same $(n-1)$-chain. If $n+2$ points, no $n+1$ of which are in the same $(n-1)$-chain, are fixed, the collineation is the identity.

§ 6. *Collineation groups in the* $PG(k, p^n)$.

The collineation group of the $PG(k, p^n)$ is defined as the group of all collineations in the $PG(k, p^n)$. In § 5, it has been proved for the $PG(k, p)$ that a collineation of a line may have at most two fixed points. As in ordinary projective geometry, it can be proved for the $PG(k, p)$ that there is a collineation determined by any three points of the line, i. e., there is one and only one collineation that transforms a given set of three points into another given set of three points. In like manner it can be proved that there is one and only one collineation of a plane determined by any four points no three of which are in the same line, and that one and only one collineation of an l-space is determined by any $l+2$ points no $l+1$ of which are in the same $(l-1)$-space.

* Cf. another definition of such a configuration in § 9.

† A circle is a 1-sphere, a sphere is a 2-sphere, etc.

Let A, B, C be three points of a line. By a collineation, A may be transformed into any point A' of the line (A' may be chosen in $p+1$ ways): B may then be transformed into any point B' other than A' (B' may be chosen in p ways); and C may be transformed into any point C' other than A' and B' (C' may chosen in $p-1$ ways). The collineation is then completely determined, and the order of the collineation group of the line is seen to be $(p+1)\,p(p-1)=p(p^2-1)$.

Let A and B be two points of a plane. Let σ_2 denote the number of points in the plane and let σ_1 denote the number of points on a line. By a collineation, A may be transformed into any point A' of the plane (A' need not be different from A and may therefore be chosen in σ_2 ways); B may be then transformed into any point B' other than A' (B' may be chosen in σ_2-1 ways); any other point C of the line AB may then be transformed into a point C' of the line $A'B'$(C' may be chosen in $p-1$ ways because $\sigma_1=p+1$): any point D not on the line AB may then be transformed into any point D' not on the line $A'B'$(D' may be chosen in $\sigma_2-\sigma_1$ ways), and finally any point E, other than A and D of the line AD, may be transformed into any point E', other than A' and D', of the line $A'D'$(E' may be chosen in $p-1$ ways). The collineation is then completely determined, and the order of the collineation group of the plane is the product of the factors which represent the number of choices at the successive stages of the determination of the collineation, i. e., the order is $\sigma_2(\sigma_2-1)(p-1)\cdot(\sigma_2-\sigma_1)(p-1)$.

In like manner it may be seen that the order of the collineation group of k-space is

$$N=\sigma_k(\sigma_k-1)(p-1)\cdot(\sigma_k-\sigma_1)(p-1)\cdot(\sigma_k-\sigma_2)(p-1)\cdots(\sigma_k-\sigma_{k-1})(p-1),$$

where $\sigma_k=p^k+p^{k-1}+\cdots+p+1=(p^{k+1}-1)/(p-1)$, the number of points in k-space ($\sigma_0=1$). The expression for N may be written

$$N=(p-1)^k\sigma_k\prod_{j=0}^{k-1}(\sigma_k-\sigma_j).$$

Substitute the values for σ_k and σ_j, as indicated above, and the expression becomes

$$N=(p-1)^k\left(\frac{p^{k+1}-1}{p-1}\right)\prod_{j=0}^{k-1}\left[\frac{p^{k+1}-p^{j+1}}{p-1}\right]\cdot$$

$$=\frac{(p^{k+1}-1)\prod_{i=1}^{k}(p^{k+1}-p^i)}{p-1}=\frac{1}{p-1}\prod_{i=0}^{k}(p^{k+1}-p^i).$$

But this is exactly the order * of the group $LF(k+1,p)$ of all linear fractional transformations on k variables having coefficients in the $GF[p]$, and having determinant not zero.

* L. E. Dickson, *Linear Groups*, p. 87.

Therefore, by observing that every transformation of the group is a collineation, it is proved that the group $LF(k+1, p)$ is the collineation group of the $PG(k, p)$.

The collineation group of the $PG(k, p^n)$, $n > 1$, is not linear.* It will be sufficient to show this in detail for the case of the plane. The equation of a line in the $PG(2, p^n)$ is

$$a_1 x_1 + a_2 x_2 + a_3 x_3 = 0. \tag{0}$$

The result of raising this equation to the power p^k is

$$a_1^{p^k} x_1^{p^k} + a_2^{p^k} x_2^{p^k} + a_3^{p^k} x_3^{p^k} = 0. \tag{k}$$

This equation contains only three terms because every other term in the expansion has a coefficient which is a multiple of p and is therefore congruent to zero modulo p. Conversely, the result of raising equation (k) to the power p^{n-k} is equation (0) because $\rho^{p^n} = \rho$ if ρ is any mark of the $GF[p^n]$. Therefore equations (0) and (k) represent the same line or, in other words, any line may be represented by any one of the n equations (k), $(k = 0, 1, \cdots, n-1)$, $n-1$ of which are not linear. Moreover, any equation of the form

$$a_1 x_1^{p^k} + a_2 x_2^{p^k} + a_3 x_3^{p^k} = 0 \tag{1}$$

represents a line since it becomes linear upon being raised to the power p^{n-k}. The transformation

$$(A_k): \quad \begin{cases} x_1' = \dfrac{a_1 x_1^{p^k} + b_1 x_2^{p^k} + c_1 x_3^{p^k}}{a_3 x_1^{p^k} + b_3 x_2^{p^k} + c_3 x_3^{p^k}}, \\ x_2' = \dfrac{a_2 x_1^{p^k} + b_2 x_2^{p^k} + c_2 x_3^{p^k}}{a_3 x_1^{p^k} + b_3 x_2^{p^k} + c_3 x_3^{p^k}}, \end{cases}$$

is a collineation because it transforms any equation of type (1) into another of type (1). It is therefore evident that the group of all collineations in the $PG(k, p^n)$, $n > 1$, is not the linear fractional group. The group of all collineations is not further considered in this paper.

§ 7. *Linear groups.*

To define synthetically, in the $PG(k, p^n)$, the group $LF(k+1, p^n)$,† $n > 1$, it is necessary to make use of the fact that condition VII, which is a valid proposition in the $PG(k, p^n)$, is equivalent to the following form of the fun-

* This is analogous to the fact that the collineation group is not linear in ordinary complex projective geometry.

† This symbol is used in this paper to denote the group of linear fractional transformations having determinant $\neq 0$ although DICKSON, in his *Linear Groups*, uses it for the group of linear fractional transformations of determinant unity.

damental theorem of projective geometry.* *If four collinear points A, B, C, D are transformed by a finite number of projections and sections into four points A_1, B_1, C_1, D_1, then, if by any finite number of projections and sections A, B, C, D are transformed into A_1, B_1, C_1, D_1', D_1' cannot be different from D_1.* This proposition should be carefully distinguished from the stronger form of the fundamental theorem which holds in the $PG(k, p)$ and in ordinary real projective geometry, namely, *Any transformation of points into points and lines into lines which transforms any set of four harmonic points into four harmonic points leaves a line l invariant point by point if three points of l are fixed.* The weak form is the one which holds in ordinary complex projective geometry.

A projective collineation is defined as a collineation such that, if four collinear points are transformed into four collinear points, the transformation can be effected by a finite number of projections and sections. The group of all such transformations is called the projective group of the $PG(k, p^n)$. The order of this group may be counted by the method used in § 6 for the collineation group of the $PG(k, p)$. The result is the same except that in the present case p^n replaces p. The order is $[1/(s-1)]\prod_{i=0}^{k}(s^{k+1}-s^i)$, where $s=p^n$. This is exactly the order of the group $LF(k+1, p^n)$† of all linear fractional transformations on k variables having coefficients in the $GF[p^n]$ and having determinant not zero.

To prove that the group $LF(k+1, p^n)$ is the projective group of the $PG(k, p^n)$ it is now sufficient to prove that every such linear fractional transformation is projective. Let A, B, C, D be four collinear points and let A, B', C', D' be four collinear points on another line through A. It is to be proved that the linear fractional transformation which transforms A, B, C, D into A, B', C', D' is a projective transformation, i. e., that the lines BB' and CC' meet on the line DD'. To prove this, the points A, D, D' (Fig. 6) are taken as the points (001), (010), (100), i. e., the triangle ADD' is taken as the fundamental triangle of the coördinate system. B and C are taken as $(01b)$ and $(01c)$, any two points of the line $x_1=0$. B' and C' are taken as $(10b')$ and $(10c')$ of the line $x_2=0$. The most general transformation of the group $LF(3, p^n)$ which transforms the line AD into the line AD' leaving the point A fixed and transforming B into B' and D into D' is easily determined to be

$$\frac{x_1'}{x_3'}=\frac{lx_1+bx_2}{nx_1+b'x_3}, \qquad \frac{x_2'}{x_3'}=\frac{mx_1}{nx_1+b'x_3},$$

l, m, n being arbitrary marks of the field. This transformation converts the

* F. Schur, Mathematische Annalen, vol. 51 (1898), p. 401. Also cf. *Notes*, p. 71.
† L. E. Dickson, *Linear Groups*, p. 87.

point $(01c)$ into the point $(1, 0, b'c/b)$, i. e., c' is determined to be $c' = b'c/b$. The equations of the lines BB' and CC' are now found to be, respectively, $b'x_1 + bx_2 - x_3 = 0$ and $b'cx_1 + bcx_2 - bx_3 = 0$. The point of intersection of

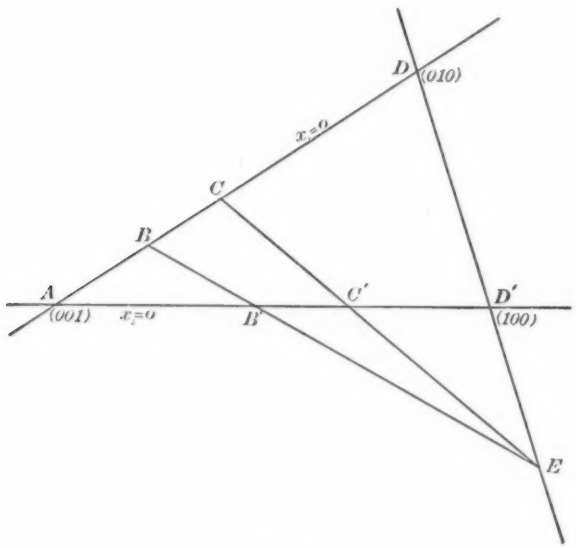


FIG. 6.

these two lines is the point E whose coördinates are $(b, b', 0)$. E is a point of the line $x_3 = 0$, i. e., of the line DD'. Since the points $ABCD$ are projective with the points A, B', C', D' from the center of projection E, the linear fractional transformation of A, B, C, D into A, B', C', D' is a projective transformation.

Having proved that the linear fractional transformation of four collinear points A, B, C, D into the four collinear points A, B', C', D' ($B'C'D'$ being any three points on any second line through A) is a projective transformation, we now observe that the linear fractional transformation of four collinear points A, B, C, D into four collinear points may be regarded as the product of two projective linear fractional transformations of the kind just considered (Fig. 7). The points A, B, C, D are first transformed into A, B', C', D_1, four points of the line AD_1, and these points are then transformed into A_1, B_1, C_1, D_1 of the given line through D_1.

It has now been proved that the group $LF(k+1, p^n)$ is the projective group of the $PG(k, p^n)$. This projective group is the sub-group (A_0) of the group (A_k) given in the last part of §6. In the $PG(k, p)$, i. e., when

$n = 1$, the group (A_k) consists only of the transformations A_0, and the projective group is the same as the collineation group.

The $PG(k, p^n)$ may now be regarded as a defining invariant for the group $LF(k+1, p^n)$, (under projective transformations of course, although when $n = 1$ this is the same as saying under collineations). If there be left off from the $PG(k, s)$, $s = p^n$, the points of a single $PG(k-1, s)$, there will remain

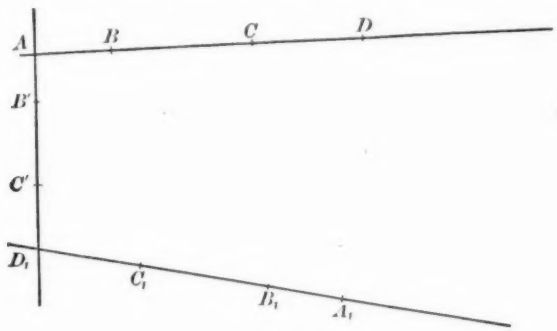


FIG. 7.

a configuration analogous to ordinary euclidean k-space. It is denoted by the symbol $EG(k, s)$. The $PG(k, s)$ may be thought of as arranged in $(s^{k+1}-1)/(s-1)$ geometries of the type $PG(k-1, s)$. The leaving off of the single $PG(k-1, s)$ takes from each remaining $PG(k-1, s)$ a single $PG(k-2, s)$. This leaves as the $EG(k, s)$ a configuration consisting of s^k points arranged in

$$\frac{s^{k+1}-1}{s-1} - 1 = \frac{s(s^k-1)}{s-1}$$

sets of s^{k-1} points each, each set being an $EG(k-1, s)$. Let $x_{k+1} = 0$ be the equation of the $PG(k-1, s)$ which was left off. The subgroup of the $LF(k+1, s)$, consisting of those transformations which leave invariant the $PG(k-1, s)$ whose equation is $x_{k+1} = 0$, is the linear group, $L(k, s)$, in k variables. It may be thought of as the group for which the $EG(k, s)$ is a defining invariant (under projective transformations, of course). The subgroup of the linear group $L(k, s)$, consisting of those transformations which leave invariant the origin of coördinates, is the general linear homogeneous group $GLH(k, s)$ in k variables. The configuration which is a defining invariant for this group is the configuration obtained by leaving off from the $EG(k, s)$ a single point and every $EG(k-1, s)$ containing it.

§ 8. *Second degree loci.*

A *point conic* is defined as the locus of the point of intersection of corresponding lines of two projective pencils of lines. A *line conic* is defined as consisting of the lines that join corresponding points of two projective ranges of points. If a tangent to a point conic be defined as a line that has one and only one point in common with the point conic, it can be proved (cf. *Notes*, p. 110), that the tangents to a point conic constitute a line conic. This being so, it is convenient to use the word *conic* to denote the self-dual figure that consists of a point conic and its tangents. The number of points of a point conic is equal to $s+1$, the number of lines in a pencil; the number of lines of a line conic is $s+1$, the number of points in a range. Therefore a conic consists of $s+1$ points and $s+1$ lines.

A point which is the intersection of two tangents is said to be an *outside point.* A point which is neither a point of the conic nor an outside point is said to be an *inside point.* The total number of outside points is $\frac{1}{2}(s^2+s)$, the number of combinations of $s+1$ tangents taken two at a time. By subtraction from s^2+s+1, the number of points in the plane, the total number of inside points is found to be $\frac{1}{2}(s^2-s)$. By the principle of duality, the number of secants (a secant being a line that meets the conic in two points) is $\frac{1}{2}(s^2+s)$ and the number of lines that do not meet the conic is $\frac{1}{2}(s^2-s)$. On any secant there are $s-1$ points which are not points of the conic. Half of these are inside points and half are outside points, because *two* tangents meet in each outside point. This corresponds to the fact that in a Galois field there are as many squares as not squares since, as is proved below, a conic may be represented by an equation of the second degree.

Let $P=0$ and $Q=0$ be the equations of two lines in abridged notation. P and Q are linear homogeneous functions of the three variables x_1, x_2, x_3. The equation $\lambda P+\mu Q=0$, λ and μ being any marks of the $GF[s]$, is the equation of a line through the point of intersection of the lines $P=0$ and $Q=0$. Therefore the equation $\lambda P+\mu Q=0$, the domain for λ and μ being the $GF[s]$, represents the pencil of lines through the point of intersection of $P=0$ and $Q=0$. There is one and only one line of the pencil for every value of the fraction λ/μ. There are s values of this fraction in the $GF[s]$. When $\mu=0$ the fraction λ/μ is denoted by the symbol ∞, and the corresponding line of the pencil is the line $P=0$.

Consider two pencils of lines $\lambda P+\mu Q=0$ and $\lambda P'+\mu Q'=0$. A one to one correspondence between the two pencils may be established by taking as corresponding elements the two lines which correspond to the same value of the parameter λ/μ. That this correspondence is projective may be proved by observing that the linear fractional transformation [a transformation of the projective group of the $PG(2, s)$] which transforms the lines $P=0$, $Q=0$

and $P + Q = 0$ into the lines $P = 0$, $Q = 0$ and $P + Q = 0$, respectively, also transforms the line $\lambda P + \mu Q = 0$ into the line $\lambda P' + \mu Q' = 0$. The locus of the point of intersection of corresponding lines is obtained by eliminating λ and μ from the equations of the two pencils. The result is $PQ' + P'Q = 0$ which is a homogeneous equation of the second degree in x_1, x_2, x_3 since P, Q, P', and Q' are linear homogeneous functions. Similarly, using line coördinates, it may be proved that the line equation of a conic is a homogeneous equation of second degree in three variables.

As in ordinary analytic or synthetic geometry, pole and polar may be defined with reference to a conic, and the conic may be thought of as consisting of those points which are such that each is contained in its corresponding line in a polar system. (For a definition of polar system without reference to a conic, cf. *Notes*, p. 147.)

A configuration analogous to the non-euclidean plane geometry of LOBATCHEWSKY may be obtained by considering only the inside points of a conic. The number of points in such a configuration is $\frac{1}{2}(s^2 - s)$. These points will be arranged in s^2 lines, the tangents being the only lines of the plane which are excluded. Every secant line of the plane contains $\frac{1}{2}(s - 1)$ inside points, and every line of the plane which does not meet the conic contains $\frac{1}{2}(s + 1)$ inside points. Of course a line joining an inside point and an outside point does not necessarily meet the conic. The configuration may be thought of as consisting of $\frac{1}{2}(s^2 - s)$ points arranged in s^2 lines which are of two classes: In the first class there are $\frac{1}{2}(s^2 + s)$ lines containing $\frac{1}{2}(s - 1)$ points each. In the second class there are $\frac{1}{2}(s^2 - s)$ lines containing $\frac{1}{2}(s + 1)$ points each.

Configurations analogous to k-dimensional non-euclidean geometries may be obtained in a similar manner from second degree loci in k dimensions.

In the geometry of three-space a quadric may be defined by means of two correlative bundles (a bundle of lines correlated with a bundle of planes). The quadric consists of the points in which a line of one bundle meets its corresponding plane in the other. This, the general definition of a quadric in projective geometry, includes the special cases of ruled quadric and quadric cone. A *ruled quadric* may be defined as the points of a set of lines each of which meets three given skew lines; or it may be thought of as the points of the lines of intersection of two projective axial pencils whose axes have no common point. A *quadric cone* is defined in like manner by means of projective axial pencils whose axes do have a common point. From either definition it appears that one ruling of a ruled quadric consists of $s + 1$ lines of $s + 1$ points each, i. e., the surface consists of $(s + 1)^2$ points. A quadric cone consists of $s + 1$ lines having a point in common, the total number of points being $s^2 + s + 1$. This is the number of points in a plane, or of lines in a bundle, or of planes in a bundle. The number of points of the general quadric is $s^2 + 1$. As in ordin-

ary projective geometry, it may be seen that any plane section of a quadric is a conic.

§ 9. *Derived tactical configurations.*

The finite projective geometries may be used to build a practically endless sequence of tactical configurations.* For example, in a three dimensional geometry, systems of lines may be obtained analogous to the ordinary linear congruence. Any three lines of the system determine a set of $p^n + 1$ lines, namely, the lines of the same ruling in the quadric determined by the three. We thus have a generalization from the case p to the case p^n of the configuration of all the chains in a line (cf. § 5).

The configurations just named seem not to be in the literature, but all the other configurations discussed above have been known previously for the cases where p^n is a simple prime. The references, so far as we have been able to find them, are given below. We have found no previous mention of a case p^n where $n > 1$.

The $PG(2, 2)$ is the same configuration as the triple system Δ_7 in seven elements mentioned in § 1 and § 3.

The $PG(k-1, 2)$ is the same configuration as the *linear triple system* Δ_{2^k-1} in $2^k - 1$ elements, defined and studied by MOORE.†

The $EG(k, 3)$ is a triple system Δ_{3^k} in 3^k elements. The number of triples is $\frac{1}{6}3^k(3^k - 1)$.

The $PG(k, p)$ is the same as the linear fractional configuration $LFC_f[(p^k - 1)/(p - 1)]$ obtained by MOORE‡ as a defining tactical invariant for the linear fractional group $LF(k+1, p)$. Paragraph 7 of this paper may be regarded as a generalization by a different method of the $LF(k+1, p)$ to the $LF(k+1, p^n)$, $n > 1$.

The $PG(k, p)$ has been defined synthetically by FANO § and analytically by HESSENBURG.‖ Neither FANO nor HESSENBERG seems to have studied its group properties or its geometrical theory. MOORE studied only the group properties. He obtained the configuration from the $GF[p^k]$ instead of from the $GF[p]$ as in this paper.

* For the general definition of tactical configuration and a rich collection of examples, cf. E. H. MOORE, *Tactical memoranda*, American Journal of Mathematics, vol. 18 (1896), p. 264.

† *Concerning the general equations of the seventh and eighth degrees*, Mathematische Annalen, vol. 51 (1898), pp. 417–444.

‡ *Concerning Jordan's linear groups*, Bulletin of the American Mathematical Society, ser. 2, vol. 2 (1895–96), pp. 33–43.

§ *Sui postulati fondamentali della geometria projettiva*, Giornale di Matematiche, vol. 30 (1892), p. 106.

‖ *Über die projective Geometrie*, Sitzungsberichte der Berliner mathematischen Gesellschaft, 1902–03, pp. 36–40.

The $EG(k, p)$ is the linear configuration $LCf[p^k]$ obtained by MOORE* as a defining tactical invariant for the linear group $L(k, p)$. The linear homogeneous configuration $LHCf(p^k - 1)$, obtained by MOORE* as a defining tactical invariant for the general linear homogeneous group $GLH(k, p)$, is the same configuration, for the case $s = p$, as the one obtained as a defining invariant for the general case of the same group at the end of § 7 of this paper.

The $EG(3, 2)$ is MOORE's quadruple system $\square_8$.† The planes of the $EG(3, 2)$ are the quadruples.

*Bulletin of the American Mathematical Society, loc. cit.

†Mathematische Annalen, loc. cit., § 5. In this paper, MOORE points out that the $PG(3, 2)$ [his Δ_{15}] leads to a solution of KIRKMAN's fifteen-schoolgirl problem.

ON THE ANALYTIC EXTENSION OF FUNCTIONS DEFINED BY DOUBLE POWER SERIES*

BY

W. B. FORD

Problem and Theorem.

1. The object of the present paper is to study the following general problem: Given the double power series

$$\sum_{m=0}^{m=\infty} \sum_{n=0}^{n=\infty} a(m, n) x^m y^n, \tag{1}$$

defining a function $f(x, y)$ of the two independent complex variables x, y within the regions $|x| < R_1$, $|y| < R_2 (R_1 > 0, R_2 > 0)$, to determine the value of this function at any point in its domain of existence.†

It is our special purpose to demonstrate the following theorem:

THEOREM. *If in the series* (1) *the function* $a(u, v)$ *considered as a function of the two independent, complex variables* u, v $(u = \alpha + i\beta, v = \gamma + i\delta)$ *is single valued and analytic for those values of* u, v *for which* $\alpha \geqq 0$, $\gamma \geqq 0$ *and if there exists a constant* c *such that for the same values of* u, v

$$|a(u, v)| < c, \tag{2}$$

* Presented to the Society February 24, 1906. Received for publication November 4, 1905.

† The importance of this problem in the theory of functions of two or more complex variables has long been recognized, but the results obtained in this field appear to be very meager. For a general statement respecting both the problem and the literature upon it see HADAMARD, *La série de Taylor et son prolongement analytique* (Paris, 1900), chap. IX, § 5. This problem is, in fact, the natural generalization of the problem of the analytic extension of a function defined by a single (instead of a double) power series (1), this latter problem being one which has been solved within recent years in a large variety of cases. Especially we may mention in this connection the results embodied in the following memoirs: *Sur les séries divergentes et les fonctions définies par un développement de Taylor*, by LE ROY in the Annales de la Faculté des Sciences de Toulouse, ser. 2, vol. 2 (1900); *Quelques applications d'une formule sommatoire générale*, by LINDELÖF in the Acta Societatis Scientiarum Fennicae, vol. 31 (1902); *Sur la fonction définie par une série de Maclaurin*, by FORD in the Journal de Mathématiques, ser. 5, vol. 9, pp. 223-232 (1903). Up to the present time the authors cited appear not to have generalized their theorems for the case of two or more variables. We propose, therefore, in the present paper to furnish such a generalization for the third of the memoirs just cited. We shall confine ourselves, for the sake of brevity, to the case of *two* variables, although the generalization for n variables can be deduced in a similar manner, i. e., by the use of CAUCHY's generalized integral theorem for functions of n variables.

then the function $f(x, y)$ defined by (1) *possesses a branch which is single valued and analytic throughout the x, y planes, exception being made of those point pairs x, y in which either x or y lies upon the rectilinear cut $0 \cdots +\infty$ in its plane. Moreover, this branch is defined by the formula*

$$f(x, y) = a(0, 0) - \tfrac{1}{2}\int_{-\infty}^{\infty} a(\tfrac{1}{2} + i\beta, 0)\,\frac{(-x)^{\frac{1}{2}+i\beta}}{\cosh \pi\beta}\,d\beta$$

$$(3) \qquad -\tfrac{1}{2}\int_{-\infty}^{\infty} a(0, \tfrac{1}{2} + i\delta)\,\frac{(-y)^{\frac{1}{2}+i\delta}}{\cosh \pi\delta}\,d\delta$$

$$+\tfrac{1}{4}\int_{-\infty}^{\infty} d\beta \int_{-\infty}^{\infty} a(\tfrac{1}{2} + i\beta, \tfrac{1}{2} + i\delta)\,\frac{(-x)^{\frac{1}{2}+i\beta}\,(-y)^{\frac{1}{2}+i\delta}}{\cosh \pi\beta \cosh \pi\delta}\,d\delta,$$

where in terms of the principal branch of the logarithm function *

$$(-x)^{\frac{1}{2}+i\beta} = e^{(\frac{1}{2}+i\beta)\log(-x)}, \qquad (-y)^{\frac{1}{2}+i\delta} = e^{(\frac{1}{2}+i\delta)\log(-y)}$$

and where, if desired, we may invert the order of integration in the improper double integral.†

Preliminary Results.

2. In order to prove this theorem we begin by observing several elementary relations to which frequent reference will be made.

First: If u and x are two complex quantities having the respective forms:

$$u = \alpha + i\beta, \qquad x = \rho e^{i\phi}, \qquad -2\pi < \phi \leqq 0,$$

then, always in terms of the principal branch * of the logarithm function,

(I) $$(-x)^u = e^{u\log(-x)} = (-x)^{\alpha} e^{-\beta(\phi+\pi)} e^{i\beta\log\rho},$$

and further

(II) $$\sin \pi u = \Omega(\alpha, \beta) \sinh \pi\beta,$$

where $\Omega(\alpha, \beta)$ is a function of α and β, such that, for all (real) values of α and β,

(III) $$|\Omega(\alpha, \beta)| \geqq 1.$$‡

* So that, for

$$x = \rho e^{i\phi}, \qquad -2\pi < \phi \leqq 0,$$

we have

$$\log(-x) = \log\rho + i(\phi + \pi).$$

† This theorem holds true under much less restrictive conditions, as appears from an examination of the analysis employed in the subsequent proof, but, for the sake of simplicity we shall confine ourselves as above. It will be observed, however, that this case covers a large class of series (1) in which $R_1 > 1$, $R_2 > 1$.

It is also desirable to remark that $f(x, y)$ as defined by (3) coincides in value with (1) whenever $|x| < R_1$, $|y| < R_2$ and neither x nor y is real and positive. For point pairs x, y in which either x or y is real and positive formula (3) in general loses significance, regardless of what $|x|$, $|y|$ may be.

‡ The proof of this and of the other relations is given below.

Again: If $\Phi(u)$ and $U(u)$ are two functions of the complex variable u, each single valued and analytic in a region A of the u plane and such that $U(u)$ vanishes in A only at the points $u = \lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_n$ which are zeros of the first order, then, for any closed contour C_n within A and surrounding the points $u = \lambda_1, \lambda_2, \cdots, \lambda_n$ we have

$$\text{(IV)} \qquad \frac{1}{2\pi i}\int_{C_n} \frac{\Phi(u)}{U(u)}\,du = \sum_1^n \frac{\Phi(\lambda_n)}{U'(\lambda_n)},$$

the indicated integration being extended over C_n in the positive direction.

Relation (IV) admits of the following generalization:

Given two complex planes P_1 and P_2 to which belong respectively the two independent complex variables u, v and given two functions $\Phi(u, v)$, $U(u, v)$ of these variables, each single valued and analytic for all values of u, v within the regions A_1, A_2 of P_1, P_2 respectively. Furthermore, suppose that the function $U(u, v)$ is of the form $U(u, v) = E(u)F(v)$, in which $E(u)$ does not vanish within A_1 except at the points $u = \lambda_1, \lambda_2, \cdots, \lambda_m$ which are zeros of the first order, and let us suppose likewise that $F(v)$ does not vanish within A_2 except at the points $v = \mu_1, \mu_2, \mu_3, \cdots, \mu_n$ which are zeros of the first order. Then, if we designate by C_m, C_n closed contours lying within A_1, A_2, respectively, and surrounding the points $u = \lambda_1, \lambda_2, \cdots, \lambda_m$; $v = \mu_1, \mu_2, \cdots, \mu_n$, we have

$$\text{(V)} \qquad \frac{1}{(2\pi i)^2}\int_{C_m} du \int_{C_n} \frac{\Phi(u, v)}{U(u, v)}\,dv = \sum_{g=1}^{g=m}\sum_{h=1}^{h=n}\left[\frac{\Phi(u, v)}{\dfrac{\partial^2 U}{\partial u\,\partial v}}\right]_{\substack{u=\lambda_g \\ v=\mu_h}},$$

the integrations being extended over C_m and C_n in the positive direction.

Formula (I) is readily seen, while the relation (II, III) appears as follows:

$$\sin \pi u = \sin(\pi\alpha + i\pi\beta) = \sin \pi\alpha \cosh \pi\beta + i \cos \pi\alpha \sinh \pi\beta = \Omega(\alpha, \beta)\sinh \pi\beta,$$

where

$$\Omega(\alpha, \beta) = \sin \pi\alpha \coth \pi\beta + i \cos \pi\alpha.$$

But, for the function $\Omega(\alpha, \beta)$ thus defined, we have

$$|\Omega(\alpha, \beta)| = \sqrt{\sin^2 \pi\alpha \coth^2 \pi\beta + \cos^2 \pi\alpha}$$

and, since for β real $|\coth \pi\beta| \geqq 1$, this last relation gives also

$$|\Omega(\alpha, \beta)| \geqq 1.$$

Relation (IV) is a direct consequence of CAUCHY's integral theorem when we observe that the residue of the function $\Phi(u)/U(u)$ at the pole $u = \lambda_n$ is $\Phi(\lambda_n)/U'(\lambda_n)$.

Similarly, relation (V) results from a well-known theorem* in the theory of

* Cf. OSGOOD, *Encyklopädie*, II B1, § 42.

functions of two or more complex variables, it being observed in the present instance that the residue of the function $\Phi(u, v)/U(u, v)$ at the pole $u = \lambda_g$, $v = \mu_h$ is

$$\left[(u-\lambda_g)(v-\mu_h)\frac{\Phi(u, v)}{U(u, v)}\right]_{\substack{u=\lambda_g \\ v=\mu_h}} = \left[\frac{\Phi(u, v)}{\frac{\partial^2 U}{\partial u \partial v}}\right]_{\substack{u=\lambda_g \\ v=\mu_h}}.$$

It is desirable to prove also the following lemma respecting the properties of double improper integrals:

Lemma. Given the function $f(\beta, t)$ of the two real variables β, t continuous throughout the region $\beta \geqq g$, $t \geqq a$ and satisfying the following conditions:

$$(a) \qquad |f(\beta, t)| \leqq \phi(\beta) \quad [\beta \geqq g, t \geqq a, \phi(\beta) \text{ continuous},]$$

$$(b) \qquad \int_g^\infty \phi(\beta)\, d\beta \text{ converges},$$

$$(c) \qquad t^\kappa |f(\beta, t)| \leqq \phi(\beta) \qquad (\beta \geqq g, t \geqq \tau \geqq a, \kappa > 1),$$

then both the double integrals

$$(4) \qquad \int_a^\infty dt \int_g^\infty f(\beta, t)\, d\beta, \qquad \int_g^\infty d\beta \int_a^\infty f(\beta, t)\, dt$$

converge and their values are equal.

Proof. In the first of the integrals (4) the single improper integral

$$(5) \qquad \int_g^\infty f(\beta, t)\, d\beta,$$

when considered for any special value of $t \geqq a$, converges absolutely, since, for such a value of t we may write as a result of (a) and (b),

$$|f(\beta, t)| \leqq \phi(\beta) \qquad (\beta \geqq g),$$

where $\phi(\beta)$ is continuous and such that the integral

$$\int_g^\infty \phi(\beta)\, d\beta$$

exists. Moreover, we shall now show that the integral (5) defines a *continuous* function $F_1(t)$ of t when $t \geqq a$ such that the limiting integral

$$(6) \qquad \int_a^\infty F_1(t)\, dt$$

exists, thus demonstrating that the first of the double integrals (4) has a meaning.

Now, the improper integral (5) will define a function of t, continuous when

$t \geqq a$, provided that for all such values of t the same integral converges *uniformly*,* i. e., an arbitrarily small positive quantity ϵ having been chosen, it shall be possible to find a positive quantity G (independent of t) such that whenever $t \geqq a$ we may write

$$\left| \int_\alpha^s f(\beta, t)\, d\beta \right| < \epsilon \qquad (s > G). \tag{7}$$

That such a relation exists appears as follows: Having chosen ϵ, let us take G, as we may do by virtue of (b), so large that

$$\int_\alpha^s \phi(\beta)\, d\beta < \epsilon \qquad (s > G). \tag{8}$$

Then availing ourselves of (a) we may write for $t \geqq a$

$$\left| \int_\alpha^s f(\beta, t)\, d\beta \right| \leqq \int_\alpha^s |f(\beta, t)|\, d\beta \leqq \int_\alpha^s \phi(\beta) d\beta < \epsilon \qquad (s > G). \tag{9}$$

But if the function

$$F_1(t) = \int_\gamma^\infty f(\beta, t)\, d\beta \tag{10}$$

is continuous when $t \geqq a$, then for any special value $r > a$ the integral

$$\int_a^r F_1(t)\, dt$$

will exist.† Moreover if we consider only those values of t for which $t \geqq \tau$ we may write, as a result of (c),

$$|F_1(t)| \leqq \int_\gamma^\infty |f(\beta, t)|\, d\beta \leqq \frac{1}{t^\kappa} \int_\gamma^\infty \phi(\beta) d\beta,$$

i. e., when $t \geqq \tau$ we shall have

$$|F_1(t)| \leqq \frac{c}{t^\kappa}, \qquad c = \int_\gamma^\infty \phi(\beta)\, d\beta \qquad (\kappa > 1). \tag{}$$

Whence, the integral (6) converges absolutely.

Having shown that the first of the integrals (4) has a meaning, we are to show that the same is true of the second of these integrals and that the two limits thus defined are equal. Moreover, these further statements will now follow, as we know from a well known theorem,‡ under the following conditions:§

* Cf. Osgood, Annals of Mathematics, ser. 2, vol. 3 (1902), p. 132 (Theorem A).

† Cf. Picard, *Traité d' analyse*, vol. 1, chap. 1.

‡ Cf. Osgood, l. c., p. 135.

§ It will be recalled that the function $f(\beta, t)$ is continuous when $t \geqq a$, $\beta \geqq g$, thus fulfilling the preliminary conditions demanded by the theorem in question.

1° The integrals

$$\int_g^x f(\beta, t)\,d\beta, \qquad \int_a^x f(\beta, t)\,dt$$

shall respectively converge uniformly throughout every fixed interval $a \leqq t \leqq b$, $g \leqq \beta \leqq h$.

2° The integral

$$\int_a^t dt \int_g^x f(\beta, t)\,d\beta$$

shall converge uniformly throughout the unlimited interval $t \geqq a$.

In other words, in order to complete the proof it suffices to establish the following three relations:

An arbitrarily small positive quantity ϵ having been chosen, it is possible to determine

(A) a positive quantity G_1 (independent of t) such that, b being any fixed quantity $> a$, we may write

$$\left| \int_{G_1}^s f(\beta, t)\,d\beta \right| < \epsilon \qquad (a \leqq t \leqq b,\ s > G_1).$$

(B) a positive quantity G_2 (independent of β) such that, h being any fixed quantity $> g$, we may write

$$\left| \int_{G_2}^r f(\beta, t)\,dt \right| < \epsilon, \qquad (g \leqq \beta \leqq h,\ r > G_2).$$

(C) a positive quantity G_3 (independent of t) such that we may write

$$\left| \int_a^t dt \int_{G_3}^s f(\beta, t)\,d\beta \right| < \epsilon \qquad (t \geqq a,\ s > G_3).$$

We proceed to establish these relations in the order here indicated.

The fact that relation (A) exists follows *a fortiori* from the existence already shown of relation (7). In fact, if relation (7) holds for the unlimited interval $t \geqq a$ it must hold for the portion $a \leqq t \leqq b$ of this interval, so that for the establishment of (A) it suffices to take $G_1 = G$ where G is defined as in (8).

As to relation (B), let us choose G_2 in the following manner: Representing by M a quantity as great as the greatest value taken by $\phi(\beta)$ when $g \leqq \beta \leqq h$,* let us choose $G_2 > \tau$ and such that for the given value of $\kappa\,(\kappa > 1)$ we shall have

$$\int_{G_2}^r \frac{dt}{t^\kappa} < \frac{\epsilon}{M} \qquad (r > G_2).$$

*Since $\phi(\beta)$ is continuous by hypothesis when $\beta \geqq g$, we may evidently take for M the upper limit of $\phi(\beta)$ for the interval $g \leqq \beta \leqq h$. For an arithmetic proof of the existence of such an upper limit, see STOLZ, *Allgemeine Arithmetik*, vol. I, chap. 9, § 1.

Then, by virtue of (c) we may write for all values of β such that $g \leqq \beta \leqq h$,

$$\left|\int_{G_2}^{r} f(\beta, t)dt\right| \leqq \int_{G_2}^{r} |f(\beta, t)|dt \leqq \phi(\beta)\int_{G_2}^{r}\frac{dt}{t^{\kappa}} < M\int_{G_2}^{r}\frac{dt}{t^{\kappa}} < \epsilon \quad (r > G_2),$$

and this suffices for the proof of the relation in question.

In order to establish relation (C) we first observe that the quantity τ may, without further restrictions in our hypotheses, be considered positive, so that the quantity $\tau^{1-\kappa}(\kappa > 1)$ may in any case be considered a positive constant. This premised, let us now choose G_3, as we may do by virtue of (b), so that

$$\int_{G_3}^{s}\phi(\beta)d\beta < \frac{\epsilon}{2c} \qquad \left(c = \tau - a + 2\frac{\tau^{1-\kappa}}{\kappa - 1},\ s > G_3\right),$$

With this choice of G_3 let us first consider only those values of t such that $a \leqq t \leqq \tau$. Utilizing (a) we may then write

$$\left|\int_a^t dt\int_{G_3}^{s} f(\beta, t)d\beta\right| \leqq \int_a^t dt\int_{G_3}^{s}|f(\beta, t)|d\beta \leqq \int_a^{\tau} dt\int_{G_3}^{s}\phi(\beta)d\beta$$

$$= (\tau - a)\int_{G_3}^{s}\phi(\beta)d\beta < c\int_{G_3}^{s}\phi(\beta)d\beta < \frac{\epsilon}{2},$$

i. e., for our choice of G_3 we have the following relation

$$\left|\int_a^t dt\int_{G_3}^{s} f(\beta, t)d\beta\right| < \frac{\epsilon}{2} < \epsilon \qquad (a \leqq t \leqq \tau,\ s > G_3). \tag{11}$$

Secondly, for those values of $t > \tau$ we now have

$$\left|\int_a^t dt\int_{G_3}^{s} f(\beta, t)d\beta\right| \leqq \left|\int_a^{\tau} dt\int_{G_3}^{s} f(\beta, t)d\beta\right| + \left|\int_{\tau}^{t} dt\int_{G_3}^{s} f(\beta, t)d\beta\right|, \tag{12}$$

in which the first term of the second member is less than $\epsilon/2$ [see (11)], while the same is true also of the second term, since by virtue of (c) we may write

$$\left|\int_{\tau}^{t} dt\int_{G_3}^{s} f(\beta, t)d\beta\right| \leqq \int_{\tau}^{t} dt\int_{G_3}^{s}|f(\beta, t)|d\beta \leqq \int_{\tau}^{t}\frac{dt}{t^{\kappa}}\int_{G_3}^{s}\phi(\beta)d\beta$$

$$= \frac{1}{\kappa - 1}(\tau^{1-\kappa} - t^{1-\kappa})\int_{G_3}^{s}\phi(\beta)d\beta < \frac{2\tau^{1-\kappa}}{\kappa - 1}\int_{G_3}^{s}\phi(\beta)d\beta < c\int_{G_3}^{s}\phi(\beta)d\beta < \frac{\epsilon}{2},$$

i. e., for our choice of G_3 relation (12) yields the following:

$$\left|\int_a^t dt\int_{G_3}^{s} f(\beta, t)d\beta\right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \qquad (t > \tau,\ s > G_3). \tag{13}$$

In summary, then, relations (11) and (13) give (C) as desired.

Proof of the Theorem.

3. The preceding relations having been noted, we proceed with the proof of the theorem stated in § 1.

For this purpose let us write the series (1) in the form

$$(14)\quad a(0,0)+\sum_1^{\infty}a(m,0)x^m+\sum_1^{\infty}a(0,n)y^n+\sum_1^{\infty}\sum_1^{\infty}a(m,n)x^m y^n.$$

Then, as a result of the hypothesis concerning the function $a(u,v)$ we may at once apply a theorem established elsewhere* to each single series here appearing. Thus the second and third terms of expression (14) may be replaced respectively by the second and third terms in the right hand member of (3), so that for the proof of the theorem of § 1 we shall here need to consider merely the function defined by the double series

$$(15)\qquad \sum_1^{\infty}\sum_1^{\infty}a(m,n)x^m y^n,$$

showing that this function is equal to the last term of (3).

For this purpose we use relation (V) of § 2, choosing in the present instance, in the notation of that relation,

$$A_1 = \textit{the positive half of the } u \textit{ plane},$$

$$A_2 = \textit{the positive half of the } v \textit{ plane},$$

$$\Phi(u,v)=\pi^2 a(u,v)(-x)^u(-y)^v,$$

$$U(u,v)=\sin\pi u\sin\pi v,$$

and taking as the contours C_m, C_n the rectangles formed by the straight lines

$$(C_m)\qquad u=\tfrac{1}{2}+i\beta,\qquad u=\alpha\pm ig,\qquad u=\tfrac{1}{2}+2\rho+i\beta,$$

$$(C_n)\qquad v=\tfrac{1}{2}+i\delta,\qquad v=\gamma\pm ih,\qquad v=\tfrac{1}{2}+2q+i\delta,$$

in which g and h are any two positive quantities, while ρ and q are any two positive integers.

As a result of the relation in question we arrive at the following equation

$$\sum_1^{2\rho}\sum_1^{2q}a(m,n)x^m y^n=-\frac{1}{4\pi^2}\int_{C_m}du\int_{C_n}\frac{\Phi(u,v)}{U(u,v)}dv$$

$$=-\tfrac{1}{4}\int_{C_m}du\int_{C_n}a(u,v)\frac{(-x)^u(-y)^v}{\sin\pi u\sin\pi v}dv;$$

or, placing

$$(16)\qquad \Psi(u,v)=-\frac{a(u,v)(-x)^u(-y)^v}{4\sin\pi u\sin\pi v},$$

*Cf. Ford, Journal de Mathématiques, ser. 5, vol. 9 (1903), p. 223.

we have

$$\sum_{1}^{2p}\sum_{1}^{2q} a(m,n)x^m y^n = \int_{C_m} du \int_{C_n} \Psi(u,v)\,dv. \tag{17}$$

We proceed to study in detail the second member of equation (17), availing ourselves of the hypotheses and of equation (16). By breaking up the integral

$$\int_{C_n} \Psi(u,v)\,dv$$

into the four component integrals which come respectively from the integration over the four sides $v=\gamma-ih$, $v=\frac{1}{2}+2q+i\delta$, $v=\gamma+ih$, $v=\frac{1}{2}+i\delta$ of the rectangle C_n, the second member of equation (17) takes the form

$$\int_{C_m}\Big(\int_{\frac{1}{2}}^{\frac{1}{2}+2q}\Psi(u,\gamma-ih)\,d\gamma + i\int_{-h}^{h}\Psi(u,\tfrac{1}{2}+2q+i\delta)\,d\delta + \int_{\frac{1}{2}+2q}^{\frac{1}{2}}\Psi(u,\gamma+ih)\,d\gamma + \int_{h}^{-h}\Psi(u,\tfrac{1}{2}+i\delta)\,d\delta\Big)\,du.$$

Similarly, by breaking up the integrations with respect to u in each of the four expressions just obtained the second member of (17) becomes expressible finally as the sum of sixteen double definite (proper) integrals, i. e., takes the form

$$\begin{aligned}
&\int_{\frac{1}{2}}^{\frac{1}{2}+2p} d\alpha \int_{\frac{1}{2}}^{\frac{1}{2}+2q} \Psi(\alpha-ig,\ \gamma-ih)\,d\gamma + i\int_{-g}^{g} d\beta \int_{\frac{1}{2}}^{\frac{1}{2}+2q} \Psi(\tfrac{1}{2}+2p+i\beta,\ \gamma+ih)\,d\gamma \\
&+\int_{\frac{1}{2}+2p}^{1} d\alpha \int_{\frac{1}{2}}^{\frac{1}{2}+2q} \Psi(\alpha+ig,\ \gamma-ih)\,d\gamma + i\int_{g}^{-g} d\beta \int_{\frac{1}{2}}^{\frac{1}{2}+2q} \Psi(\tfrac{1}{2}+i\beta,\ \gamma-ih)\,d\gamma \\
&+i\int_{\frac{1}{2}}^{\frac{1}{2}+2p} d\alpha \int_{-h}^{h} \Psi(\alpha-ig,\ \tfrac{1}{2}+2q+i\delta)\,d\delta - \int_{-g}^{g} d\beta \int_{-h}^{h} \Psi(\tfrac{1}{2}+2p+i\beta,\ \tfrac{1}{2}+2q+i\delta)\,d\delta \\
&+i\int_{\frac{1}{2}+2p}^{1} d\alpha \int_{-h}^{h} \Psi(\alpha+ig,\ \tfrac{1}{2}+2q+i\delta)\,d\delta - \int_{g}^{-g} d\beta \int_{-h}^{h} \Psi(\tfrac{1}{2}+i\beta,\ \tfrac{1}{2}+2q+i\delta)\,d\delta \\
&+\int_{\frac{1}{2}}^{\frac{1}{2}+2p} d\alpha \int_{\frac{1}{2}+2q}^{\frac{1}{2}} \Psi(\alpha-ig,\ \gamma+ih)\,d\gamma + i\int_{-g}^{g} d\beta \int_{\frac{1}{2}+2q}^{\frac{1}{2}} \Psi(\tfrac{1}{2}+2p+i\beta,\ \gamma+ih)\,d\gamma \\
&+\int_{\frac{1}{2}+2p}^{1} d\alpha \int_{\frac{1}{2}+2q}^{\frac{1}{2}} \Psi(\alpha+ig,\ \gamma+ih)\,d\gamma + i\int_{g}^{-g} d\beta \int_{\frac{1}{2}+2q}^{\frac{1}{2}} \Psi(\tfrac{1}{2}+i\beta,\ \gamma+ih)\,d\gamma \\
&+i\int_{\frac{1}{2}}^{\frac{1}{2}+2p} d\alpha \int_{h}^{-h} \Psi(\alpha+ig,\ \tfrac{1}{2}+i\delta)\,d\delta - \int_{-g}^{g} d\beta \int_{h}^{-h} \Psi(\tfrac{1}{2}+2p+i\beta,\ \tfrac{1}{2}+i\delta)\,d\delta \\
&+i\int_{\frac{1}{2}}^{\frac{1}{2}+2p} d\alpha \int_{h}^{-h} \Psi(\alpha-ig,\ \tfrac{1}{2}+i\delta)\,d\delta - \int_{g}^{-g} d\beta \int_{h}^{-h} \Psi(\tfrac{1}{2}+i\beta,\ \tfrac{1}{2}+i\delta)\,d\delta.
\end{aligned} \tag{18}$$

We shall now examine each of these double integrals in detail.

From the definition of $\Psi(u, v)$ given by (16), together with relation (I) of §1, we have for all (real) values of $\alpha, \beta, \gamma, \delta (\alpha > 0, \gamma > 0)$.

$$(19)\quad \Psi(\alpha + i\beta, \gamma + i\delta) = (-x)^{\alpha}(-y)^{\gamma} a(\alpha + i\beta, \gamma + i\delta) \frac{e^{i\beta[\log\rho + i(\phi+\pi)]} e^{i\delta[\log\sigma + i(\psi+\pi)]}}{-4\sin\pi(\alpha + i\beta)\sin\pi(\gamma + i\delta)},$$

where $\phi = \arg x$, $\psi = \arg y$, $\rho = \operatorname{mod} x$, $\sigma = \operatorname{mod} y$. Whence, by use of (II), the first term of (18) becomes

$$I_1 = \frac{e^{g(\phi+\pi)} e^{-ig\log\rho} e^{h(\psi+\pi)} e^{-ih\log\sigma}}{-4\sinh\pi g \sinh\pi h} \int_{\frac{1}{2}}^{\frac{1}{2}+2p} \frac{(-x)^{\alpha} d\alpha}{\Omega(\alpha, -g)} \int_{\frac{1}{2}}^{\frac{1}{2}+2q} \frac{a(\alpha - ig, \gamma - ih)(-y)^{\gamma}}{\Omega(\gamma, -h)} d\gamma.$$

Again, using (2) and (III), we now obtain

$$|I_1| < c\rho^{\frac{1}{2}+2p}\sigma^{\frac{1}{2}+2q} \frac{e^{g(\phi+\pi)} e^{h(\psi+\pi)}}{4\sinh\pi g \sinh\pi h} \int_{\frac{1}{2}}^{\frac{1}{2}+2p} d\alpha \int_{\frac{1}{2}}^{\frac{1}{2}+2q} d\gamma.$$

Whence, if neither x nor y is real and positive, we have but to recall that our values of $\phi = \arg x$ and $\psi = \arg y$ are always taken such that for the given point x, y we have $\phi < 0$, $\psi < 0$ in order to write for the point x, y, in question

$$\lim_{g=\infty,\, h=\infty} I_1 = 0.$$

Similarly, if we bear in mind throughout the restrictions

$$(20)\qquad -2\pi < \phi < 0, \qquad -2\pi < \psi < 0,$$

it appears by means of (II), (III) and (2) that the third, ninth and eleventh terms of expression (18) approach the limit zero for the given values of x, y when $g = \infty$, $h = \infty$.

We pass on, then, to consider the second term of (18). By use of (19) and the relation

$$(21)\quad \sin\pi(\tfrac{1}{2} + 2n + i\beta) = \sin\left(\frac{\pi}{2} + 2n\pi + i\pi\beta\right) = \cos i\pi\beta = \cosh\pi\beta \qquad (n = 0, 1, 2, 3, \cdots),$$

this term becomes

$$I_2 = \frac{e^{h(\psi+\pi)} e^{-ih\log\rho}(-x)^{\frac{1}{2}+2p}}{-4i\sinh\pi h} \int_{-g}^{g} \frac{e^{i\beta[\log\rho + i(\phi+\pi)]}}{\cosh\pi\beta} d\beta \int_{\frac{1}{2}}^{\frac{1}{2}+2q} a(\tfrac{1}{2} + 2p + i\beta, \gamma - ih) \frac{(-y)^{\gamma}}{\Omega(\gamma, -h)} d\gamma.$$

Whence, again using (III) and (2), we may write

$$|I_2| < c\rho^{\frac{1}{2}+2p}\sigma^{\frac{1}{2}+2q} \frac{e^{h(\psi+\pi)}}{4\sinh\pi h} \int_{-g}^{g} \frac{e^{-(\phi+\pi)\beta}}{\cosh\pi\beta} d\beta \int_{\frac{1}{2}}^{\frac{1}{2}+2q} d\gamma,$$

so that for our point x, y we have as before

$$\lim_{y=\infty,\, h=\infty} I_2 = 0.^*$$

Similarly, the same fact appears for the fourth, fifth, seventh, tenth, twelfth, thirteenth and fifteenth terms of (18).

If then we confine ourselves to values x, y such as indicated, we may replace expression (18) by the sum of those of its terms which we have not as yet considered in detail — viz.: the sixth, eighth, fourteenth and sixteenth — it being understood, however, that we now take as limits of integration in these terms $y=\infty$, $h=\infty$. In other words, expression (18) may be written in the following form involving improper double integrals:

$$-\int_{-x}^{x} d\beta \int_{-\infty}^{\infty} \Psi(\tfrac{1}{2}+2\rho+i\beta,\ \tfrac{1}{2}+2q+i\delta)d\delta + \int_{-x}^{x} d\beta \int_{-\infty}^{\infty} \Psi(\tfrac{1}{2}+i\beta,\ \tfrac{1}{2}+2q+i\delta)d\delta$$

(22)

$$+\int_{-x}^{x} d\beta \int_{-\infty}^{x} \Psi(\tfrac{1}{2}+2\rho+i\beta,\ \tfrac{1}{2}+i\delta)d\delta - \int_{-\infty}^{x} d\beta \int_{-\infty}^{\infty} \Psi(\tfrac{1}{2}+i\beta,\ \tfrac{1}{2}+i\delta)d\delta.$$

As before we shall now examine each of these double integrals in detail.

In the first place, in order to be assured that the first term of (22) has a meaning we should show† that each of the terms of the following expression:

$$-\int_{0}^{\infty} d\beta \int_{0}^{\infty} \Psi(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta)d\delta - \int_{-\infty}^{0} d\beta \int_{0}^{\infty} \Psi(\tfrac{1}{2}+2\rho+i\beta,\ \tfrac{1}{2}+2q+i\delta)d\delta$$

(23)

$$-\int_{0}^{x} d\beta \int_{-\infty}^{0} \Psi(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta)d\delta - \int_{-\infty}^{0} d\beta \int_{-\infty}^{0} \Psi(\tfrac{1}{2}+2\rho+i\beta,\ \tfrac{1}{2}+2q+i\delta)d\delta,$$

has a meaning

Confining our attention at first to the first term of (23) we shall now show that in case the usual conditions (20) are satisfied by x, y this term has a meaning, thus showing that the same term defines a function of x, y throughout the entire x, y planes, with the exception of points on the positive real axes of x and y. Moreover, we shall establish this by means of the lemma of § 2, so that it will also result from our investigations that the order of integration may be inverted in this integral if desired.

* It will be recalled that owing to relations (20) the expression

$$\lim_{y=\infty} \int_{-y}^{0} \frac{e^{-(\phi+\pi)\beta}}{\cosh \pi\beta}\, d\beta = \int_{-\infty}^{0} \frac{e^{-(\phi+\pi)\beta}}{\cosh \pi\beta}\, d\beta + \int_{0}^{\infty} \frac{e^{-(\phi+\pi)\beta}}{\cosh \pi\beta}\, d\beta$$

has a meaning.

† Cf. STOLZ, *Differential- und Integralrechnung*, vol. I, chap. 10, § 9.

According to the lemma of § 2 we may say that the first term of (23) will have a meaning under the following conditions: *

$$(a) \quad |\Psi(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta)| \leqq L(\beta) \quad [\beta \geqq 0, \delta \geqq 0, L(\beta) \text{ continuous}],$$

$$(b) \quad \int_0^\infty L(\beta)\,d\beta \text{ converges},$$

$$(c) \quad \delta^\kappa |\Psi(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta)| \leqq L(\beta) \quad (\beta \geqq 0, \delta \geqq \delta_1 \geqq 0, \kappa > 1).$$

Now from (19) and (21) we have

$$(24) \quad \Psi(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta) = -a(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta)\frac{(-x)^{\frac{1}{2}+2\rho}(-y)^{\frac{1}{2}+2q} e^{i\beta[\log\rho+i(\phi+\pi)]} e^{i\delta[\log\sigma+i(\psi+\pi)]}}{4\cosh\pi\beta\cosh\pi\delta}.$$

Whence, using (2), we obtain for all values (real) of β and δ and for any set of values for ρ and q taken from the sequences $\rho=0, 1, 2, 3, \cdots$; $q=0, 1, 2, 3, \cdots$ the relation

$$(25) \quad |\Psi(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta)| < \frac{c}{4}\rho^{\frac{1}{2}+2\rho}\sigma^{\frac{1}{2}+2q}\frac{e^{-(\phi+\pi)\beta}e^{-(\psi+\pi)\delta}}{\cosh\pi\beta\cosh\pi\delta}.$$

But, if M_ψ represents a positive quantity as large as the largest value taken by $e^{-(\psi+\pi)\delta}/\cosh\pi\delta\,(\psi > -2\pi)$ when $\delta \geqq 0$ † we obtain from (25) the relation

$$|\Psi(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta)| < L(\beta) = M_\psi \frac{c\rho^{\frac{1}{2}+2\rho}\sigma^{\frac{1}{2}+2q}}{4}\frac{e^{-(\phi+\pi)\beta}}{\cosh\pi\beta} \quad (\beta \geqq 0, \delta \geqq 0)$$

and thus condition (a) becomes satisfied. Moreover, by virtue of the relation $\phi > -2\pi$ condition (b) becomes satisfied also. As to condition (c), if we represent by δ_1 a positive quantity so large that for $\delta \geqq \delta_1$ we have

$$\delta^2\frac{e^{-(\psi+\pi)\delta}}{\cosh\pi\delta} < M_\psi \quad (M_\psi \text{ defined as above})$$

we obtain from (25) the relation

$$\delta^2|\Psi(\tfrac{1}{2}+2\rho+i\beta, \tfrac{1}{2}+2q+i\delta)| < M_\psi\frac{c\rho^{\frac{1}{2}+2\rho}\sigma^{\frac{1}{2}+2q}}{4}\frac{e^{-(\phi+\pi)\beta}}{\cosh\pi\beta} = L(\beta) \quad (\beta \geqq 0, \delta \geqq \delta_1),$$

so that this condition is also satisfied by virtue of the relation $\psi > -2\pi$.

Similarly, by availing ourselves of the lemma of § 2 and of the relations (20)

* We observe that the function $\Psi(\frac{1}{2}+2\rho+i\beta, \frac{1}{2}+2q+i\delta)$ is continuous when $\beta \geqq 0$, $\delta \geqq 0$ [see (19)]. Also, it is to be understood throughout the discussion immediately following that fixed values have been assigned to x, y, p and q; hence also to ρ, σ, ϕ and ψ.

† It will be recalled here that $\cosh\pi\delta = \frac{1}{2}(e^{\pi\delta}+e^{-\pi\delta})$ so that for the given value of ψ the number M_ψ certainly exists.

and (25) we see that conditions (a), (b) and (c) are satisfied by each of the remaining three terms of (23).

The first term of (22) therefore has a meaning for the point x, y in question. Moreover, the same can now be said of the remaining three terms of (22), since the reasoning which we have already given holds for any values of p and q taken from the sequences $p = 0, 1, 2, \cdots$; $q = 0, 1, 2, \cdots$, and the second, third and fourth terms of (22) correspond respectively to the sets $(p = 0, q = q)$, $(p = p, q = 0)$, $(p = 0, q = 0)$.

The existence of the four terms of expression (22) having been shown, it also follows directly from relation (25) that if we make the further restrictions $|x| = \rho < 1$, $|y| = \sigma < 1$, then the first term of (22) will approach the limit zero when either

$$(26) \qquad \begin{cases} 1) & p \textit{ increases indefinitely, } q \textit{ remaining constant;} \\ 2) & q \textit{ increases indefinitely, } p \textit{ remaining constant;} \\ 3) & p \textit{ and } q \textit{ increase indefinitely in any manner.} \end{cases}$$

In fact, writing as before this term in the form (23), we see from (25) that the first term of (23) is less in absolute value than

$$(27) \qquad \frac{c}{4}\rho^{\frac{1}{2}+2p}\sigma^{\frac{1}{2}+2q}\int_0^\infty \frac{e^{-(\phi+\pi)\beta}}{\cosh \pi\beta}\,d\beta \int_0^\infty \frac{e^{-(\psi+\pi)\delta}}{\cosh \pi\delta}\,d\delta,$$

in which each improper integral has a meaning by virtue of relations (20). Moreover, since $\rho < 1$, $\sigma < 1$, the expression (27) is seen to approach zero when either one of the conditions (26) is satisfied, whence the same is true of the term in question, i. e., the first term of (23). Likewise, the same properties are seen to be possessed by the second, third and fourth terms of (23) and hence by the first term of (22).

Moreover, since the second term of (22) is the special case arising from its first term when p has the constant value zero, it follows from what we have just seen that under the same conditions for x, y this second term approaches the limit zero when $q = \infty$. Likewise, the third term of (22) which is obtained from its first term by placing therein $q = constant = 0$ approaches the limit zero when $p = \infty$.

In summary then we may write, whenever $|x| < 1$, $|y| < 1$ (positive real values excluded),

$$\sum_1^\infty \sum_1^\infty a(m, n)x^m y^n = -\int_{-\infty}^{\infty} d\beta \int_{-\infty}^{\infty} \Psi(\tfrac{1}{2} + i\beta, \tfrac{1}{2} + i\delta)\,d\delta = I(x, y),$$

where

$$(28) \qquad I(x, y) = \tfrac{1}{4}\int_{-\infty}^{\infty} d\beta \int_{-\infty}^{\infty} a(\tfrac{1}{2} + i\beta, \tfrac{1}{2} + i\delta)\,\frac{(-x)^{\frac{1}{2}+i\beta}(-y)^{\frac{1}{2}+i\delta}}{\cosh \pi\beta \cosh \pi\delta}\,d\delta.$$

Moreover, we have now shown that this expression (28) has a meaning for all

values of x, y, provided merely that we choose $\phi = \arg x$, $\psi = \arg y$ so that relations (20) exist, and that we may, if desired, invert the order of integration.

We proceed to show that the function $I(x, y)$ thus defined not only has a meaning throughout any region of the x, y planes in which relations (20) exist, but that in these larger regions it is a *single valued analytic* function of the independent variables x, y thus completing the proof of the theorem in question.

That the function $I(x, y)$ is singled valued within such a region follows from the fact that the functions [see (I)]

$$(-x)^{\frac{1}{2}+i\beta} = \sqrt{-x}e^{-\beta(\phi+\pi)}e^{i\beta\log\rho}, \qquad (-y)^{\frac{1}{2}+i\delta} = \sqrt{-y}e^{-\delta(\psi+\pi)}e^{i\delta\log\sigma} \quad (\beta, \delta \text{ real})$$

are single valued in the region.

In order to prove the analytic nature of $I(x, y)$ we begin by writing

$$I(x, y) = -\int_0^\infty d\beta \int_0^\infty \Psi(\tfrac{1}{2}+i\beta, \tfrac{1}{2}+i\delta)d\delta - \int_{-\infty}^0 d\beta \int_0^\infty \Psi(\tfrac{1}{2}+i\beta, \tfrac{1}{2}+i\delta)d\delta$$

$$(29) \qquad - \int_0^\infty d\beta \int_{-\infty}^0 \Psi(\tfrac{1}{2}+i\beta, \tfrac{1}{2}+i\delta)d\delta - \int_{-\infty}^0 d\beta \int_{-\infty}^0 \Psi(\tfrac{1}{2}+i\beta, \tfrac{1}{2}+i\delta)d\delta.$$

Turning to the first term of this expression let us write

$$(30) \qquad \int_0^\infty d\beta \int_0^\infty \Psi(\tfrac{1}{2}+i\beta, \tfrac{1}{2}+i\delta)d\delta = \int_0^\infty F_1(\beta, x, y)d\beta,$$

where

$$(31) \qquad F_1(\beta, x, y) = \int_0^\infty \Psi(\tfrac{1}{2}+i\beta, \tfrac{1}{2}+i\delta)d\delta.$$

Now, in order that expression (30) shall represent a function of x, y analytic in the region mentioned above it suffices to show

1) that for any special value of $\beta \geqq 0$ the integral (31) converges *uniformly* for all values of x, y lying within two arbitrary regions T_1, T_2 of the x, y planes respectively, which regions do not cut (or touch) the positive portions of the real axes of x, y.

2) that for the same values of x, y the integral

$$(32) \qquad \int_0^\infty F_1(\beta, x, y)d\beta.$$

also converges *uniformly*.

In fact, under these circumstances the function $F_1(\beta, x, y)$, $(\beta \geqq 0)$ will be analytic throughout T_1, T_2 and likewise the same will be true of the function (32) or (30)*.

* Cf. OSGOOD, *Encyklopädie*, vol. 2, B 1, § 43.

Now, for any special point x, y in T_1, T_2 we have by placing $p = q = 0$ in (25)

$$(33) \qquad |\Psi(\tfrac{1}{2} + i\beta, \tfrac{1}{2} + i\delta)| < \frac{c}{4} \rho_0^{\frac{1}{2}} \sigma_0^{\frac{1}{2}} \frac{e^{-(\phi+\pi)\beta}\, e^{-(\psi+\pi)\delta}}{\cosh \pi\beta \cosh \pi\delta} \qquad (\beta, \delta \text{ real}).$$

Moreover, relations (20) are satisfied for every such point x, y throughout T_1, T_2 inclusive of boundary values.

Whence, if we represent by ρ_0, σ_0 respectively, the greatest values taken by ρ, σ, upon the boundaries of T_1, T_2 and by $\bar{\phi}$ and $\bar{\psi}$, respectively, the least values taken by ϕ and ψ upon the same boundaries, we may write [see (33)] for all points x, y in T_1, T_2

$$(34) \quad |\Psi(\tfrac{1}{2} + i\beta, \tfrac{1}{2} + i\delta)| < \frac{c}{4} \sqrt{\rho_0 \sigma_0} \frac{e^{-(\bar{\phi}+\pi)\beta}\, e^{-(\bar{\psi}+\pi)\delta}}{\cosh \pi\beta \cosh \pi\delta}, \quad \begin{pmatrix} \beta \geqq 0, -2\pi < \bar{\phi} < 0 \\ \delta \geqq 0, -2\pi < \bar{\psi} < 0 \end{pmatrix}.$$

Whence we see that having chosen an arbitrarily small positive quantity ϵ, we may determine a quantity G_1 (dependent upon ϵ and β), so large that for any special value of $\beta \geqq 0$ we may write for all points x, y in T_1, T_2.

$$\int_{G_1}^{s} |\Psi(\tfrac{1}{2} + i\beta, \tfrac{1}{2} + i\delta)|\, d\delta < \frac{c}{4} \sqrt{\rho_0 \sigma_0} \frac{e^{-(\bar{\phi}+\pi)\beta}}{\cosh \pi\beta} \int_{G_1}^{s} \frac{e^{-(\bar{\psi}+\pi)\delta}}{\cosh \pi\delta} d\delta < \epsilon \qquad (s > G_1),$$

i. e., condition 1) becomes satisfied by the first term of (29).

Moreover, from (34) we now have for the same points x, y

$$|F_1(\beta, x, y)| \leqq \int_0^x |\Psi(\tfrac{1}{2} + i\beta, \tfrac{1}{2} + i\delta)|\, d\delta < \frac{c}{4} \sqrt{\rho_0 \sigma_0} \frac{e^{-(\bar{\phi}+\pi)\beta}}{\cosh \pi\beta} \int_0^{\infty} \frac{e^{-(\bar{\psi}+\pi)\delta}}{\cosh \pi\delta} d\delta.$$

Whence, for any positive quantity G_2, we may write under the same conditions for x, y

$$(35) \quad \int_{G_2}^{r} F_1(\beta, x, y) d\beta \Big| < \frac{c}{4} \sqrt{\rho_0 \sigma_0} \int_{G_2}^{r} \frac{e^{-(\bar{\phi}+\pi)\beta}}{\cosh \pi\beta} d\beta \int_0^x \frac{e^{-(\bar{\psi}+\pi)\delta}}{\cosh \pi\delta} d\delta \quad (r > G_2).$$

But, since relations (20) hold in particular when $\phi = \phi$, while the improper integral here appearing has a meaning by virtue of the relation $\psi < -2\pi$, we may take G_2 so large that the second member of (35) will become (and remain) less than any preassigned positive quantity η (independent of x and y); i. e., condition 2° becomes satisfied by the first term of (29).

Thus, the first term of expression (29) has the analytic properties desired and, by availing ourselves again of relations (33) and (20), the same result is seen to be true of the three remaining terms of (29) and therefore of the expression $I(x, y)$ defined in (28).

ANN ARBOR,
November, 1905.

ON QUADRATIC, HERMITIAN AND BILINEAR FORMS*

BY

LEONARD EUGENE DICKSON

Introduction.

Part I is concerned with the reduction of quadratic forms in an arbitrary field to canonical types, a problem hitherto treated only for finite fields and for the field of all real or all complex numbers.

Part II treats of the reduction of hermitian forms in a field Q obtained by the adjunction to an arbitrary field F of a root of a quadratic equation belonging to and irreducible in F. The problem is completely solved when F is any finite field, the field of all real numbers, or the field of all rational numbers.

Part III deals with the bilinear forms in an arbitrary field F which are invariant under a given substitution S with coefficients in F. The necessary and sufficient conditions on S for the existence of such invariants are obtained, and the reduction of the invariants to a single normal form is effected by a transformation commutative with S. Here and in Part IV use is made of the writer's determination of the canonical form of a linear transformation in an arbitrary field.†

Part IV treats the analogous questions on quadratic forms and gives the generalization to an arbitrary field F of JORDAN's recent results for the case of a field of order p, a prime.‡ The explicit form of the general invariant is determined with the same ease, but the difficult problem of its reduction to canonical forms by substitutions commutative with S becomes much more troublesome for an arbitrary field than for a finite field. When F does not have modulus 2, the general character of the result may be indicated as follows. As seminormal forms of the invariant we obtain

$$\sum B_i + \sum a_i H_i + \sum a_i' H_i' + \cdots + \sum A_i + \sum b_i Q_i + \sum b_i' Q_i' + \cdots,$$

where the invariants B_i, H_i, A_i and Q_i have respectively the character of bilinear, hermitian, alternate bilinear and quadratic forms, each with completely

* Presented to the Society (Chicago) December 30, 1905. Received for publication November 26, 1905.

† American Journal of Mathematics, vol. 24 (1902), pp. 101–108.

‡ *Mémoire sur les formes quadratiques, suivant un module premier p, invariantes par une substitution linéaire donnée*, Journal de Mathématiques, ser. 6, vol. 1 (1905), pp. 217–284.

fixed coefficients, while b_i, b_i', $\cdots$ are any non-vanishing elements of F and a_j, a_j', $\cdots$, any non-vanishing elements of certain fields $F(\rho)$. The question of ultimate normal forms is the question of the extent to which these parameters can be specialized by the application of a substitution commutative with S. It is shown that normalization must take place in each sum separately, that the normalization of the quadratic function $\sum b_\alpha Q_\alpha$ of certain variables x_j^α must be made by a substitution T on the x_0^α, the same substitution on the x_1^α, the same on x_2^α, etc., with analogous remarks on the sums $\sum a_i H_i$, $\cdots$. But the effect on the b_i is the same as if we had applied to $\sum b_\alpha (x_0^\alpha)^2$ the substitution T on the variables x_0^α alone. Moreover, this partial substitution on the x_0^α may be chosen arbitrarily. Hence the problem of normalizing $\sum_{\alpha=1}^{l} b_\alpha Q_\alpha$ by a substitution commutative with S and cogredient in the various sets of variables is replaced by the problem (treated in Part I) of normalizing an l-ary quadratic form in F by means of unrestricted l-ary substitutions in F. Similarly, the problem on $\sum a_i H_i$ reduces to that on hermitian forms (Part II).

I. *Reduction of quadratic forms* in a general field F.*

1. Within any field F, not having modulus 2, a quadratic form of non-vanishing determinant is linearly reducible to †

$$q \equiv \sum_{i=1}^{n} a_i x_i^2 \qquad (\text{each } a_i \text{ an element} \neq 0 \text{ of } F). \tag{1}$$

Hence for the field of real numbers the canonical types are

$$f_s \equiv \sum_{i=1}^{s} x_i^2 - \sum_{i=s+1}^{n} x_i^2.$$

For $s \neq \sigma$, f_s cannot be transformed into f_σ by a real n-ary linear substitution; this invariance of s is the JACOBI-SYLVESTER law of inertia of real quadratic forms. ‡

2. Under the transformation

$$x_i = \sum_{j=1}^{n} b_{ij} y_j \qquad (i = 1, \cdots, n),$$

q becomes

$$\sum_{j=1}^{n} A_j y_j^2 + 2 \sum_{j,k;\, j<k}^{1,\ldots,n} B_{jk} y_j y_k, \qquad A_j \equiv \sum_{i=1}^{n} a_i b_{ij}^2, \qquad B_{jk} \equiv \sum_{i=1}^{n} a_i b_{ij} b_{ik}.$$

We discuss the question: Given $b_{11}, b_{21}, \cdots, b_{n1}$, in the general field F, such that $A_1 \neq 0$, can we determine elements $b_{ij} (j > 1)$ of F, such that every

* The same treatment applies to the reduction of symmetric bilinear forms by cogredient transformations of the two sets of variables.

† The usual proof for the field of all real numbers is valid for F. Or we may proceed as in § 6, identifying every element with its conjugate.

‡ References in BALTZER, *Determinanten*, 5th ed., p. 176.

$B_{jk} = 0$ and $\Delta \equiv |b_{ij}| \neq 0$? Since the b_{i1} enter the question symmetrically, we may assume that $b_{11} \neq 0$. The conditions $B_{1k} = 0$ are satisfied if we take

$$b_{1k} = -a_1^{-1} b_{11}^{-1} \sum_{i=2}^{n} a_i b_{i1} b_{ik} \qquad (k = 2, \cdots, n).$$

If, with these values inserted in Δ, we remove the factor $a_1^{-1} b_{11}^{-1}$ from the first row of Δ, then multiply the ith row by $a_i b_{i1}$ and add the sum to the first row, for $i = 2, \cdots, n$, we find that

$$\Delta = a_1^{-1} b_{11}^{-1} A_1 \Delta_{11}, \qquad \Delta_{11} \equiv |b_{is}| \qquad (i, s = 2, \cdots, n).$$

If $n = 2$, we take $b_{22} \neq 0$ and obtain an affirmative answer to our question. Let next $n > 2$. In $B_{jk} = 0 (2 \leqq j < k)$, we replace b_{1j} and b_{1k} by their values and get

$$B'_{jk} \equiv \sum_{i=2}^{n} R_i b_{ij} b_{ik} + \sum_{i,\, t,\, i \neq t}^{2, \cdots, n} P_{it} b_{ij} b_{tk} = 0, \quad R_i \equiv a_i^2 b_{i1}^2 + a_1 a_i b_{11}^2, \quad P_{it} \equiv a_i a_t b_{i1} b_{t1}.$$

(i) Suppose first that every $R_i = 0$ $(i = 2, \cdots, n)$. Then every $b_{i1} \neq 0$, $P_{it} \neq 0$. With the exception of b_{32}, we take every $b_{is} = 0$ $(i, s = 2, \cdots, n;\ i > s)$. We assign any value not zero to b_{ii} $(i > 1)$; and any values to b_{23}, b_{32} such that $b_{23} b_{32} = -b_{22} b_{33}$. Then $\Delta_{11} = 2 b_{22} b_{33} \cdots b_{nn} \neq 0$. Finally, for $l = 4, \cdots, n$, we set

$$C_{il} \equiv \sum_{t \neq i}^{t = 2, \cdots, l} P_{it} b_{tl} = 0 \qquad (i = 2, \cdots, l-1).$$

Then conditions $B'_{jk} = 0$ $(j, k = 2, \cdots, n;\ j < k)$ are all satisfied, since the coefficient of b_{ij} equals C_{ik} if $k \geqq 4$, while for $k = 3$, $j = 2$, the condition reduces to $P_{23}(b_{22} b_{33} + b_{32} b_{23}) = 0$. But for a fixed l, $l > 3$, equations $C_{il} = 0$ serve to express the b_{it} $(t = 2, \cdots, l-1)$ in terms of b_{1l}, since the determinant of the coefficients of the former is*

$$\begin{vmatrix} 0 & P_{23} & P_{24} & \cdots & P_{2\,l-1} \\ P_{32} & 0 & P_{34} & \cdots & P_{3\,l-1} \\ P_{42} & P_{43} & 0 & \cdots & P_{4\,l-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{l-1\,2} & P_{l-1\,3} & P_{l-1\,4} & \cdots & 0 \end{vmatrix} = r \begin{vmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & 0 \end{vmatrix} = (-1)^{l-3}(l-3)\, r,$$

* To reach the second determinant, replace each P_{it} by $a_i a_t b_{i1} b_{t1}$, remove the factor $a_i b_{i1}$ from the $(i-1)$th row, and the factor $a_t b_{t1}$ from the $(t-1)$th column. To evaluate the second determinant, subtract the first row from each of the remaining rows, then add the 2^d, $\cdots$, $(l-2)$th columns to the first. The resulting determinant has the first row $l-3, 1, 1, \cdots, 1$, and zeros elsewhere outside the main diagonal.

where $r \equiv a_2^2 a_3^2 \cdots a_{l-1}^2 b_{21}^2 b_{31}^2 \cdots b_{l-11}^2 \neq 0$. Hence the equations $C_{it} = 0$ can all be satisfied if * F does not have a modulus $\leqq n-3$.

(ii) If not every R_i vanishes, we may set $R_2 \neq 0$ in view of the symmetry. We determine the b_{2k} to make the coefficient of b_{2j} in B'_{jk} vanish:

$$b_{2k} = -R_2^{-1} \sum_{t=3}^{n} P_{2t} b_{tk} \qquad (k = 3, \cdots, n).$$

We give to b_{22} any value $\neq 0$ and set $b_{i2} = 0$ $(i = 3, \cdots, n)$. Then $B'_{2k} \equiv 0$ and

$$\Delta_{11} = b_{22} \Delta_{22}, \qquad \Delta_{22} \equiv |b_{is}| \qquad (i, s = 3, \cdots, n).$$

If $n = 3$, we take $b_{33} \neq 0$ and obtain an affirmative answer to our question. Let next $n > 3$. In $B'_{jk} = 0$ $(3 \leqq j < k)$, we replace b_{tk} for $t = 2$ by its value and obtain

$$B''_{jk} \equiv \sum_{i=3}^{n} b_{ij} \left\{ R'_i b_{ik} + \sum_{t \neq i}^{t=3, \cdots, n} P'_{it} b_{tk} \right\} = 0, \quad R'_i \equiv R_2 R_i - P_{2i}^2, \quad P'_{it} \equiv a_1 a_2 b_{11}^2 P_{it},$$

the value of P'_{it} being initially $R_2 P_{it} - P_{2i} P_{2t}$. The present problem — to determine the b_{is} $(i \geqq 3, s \geqq 3)$ so that their determinant $\Delta_{22} \neq 0$ and each $B''_{jk} = 0$ $(3 \leqq j < k)$ — is the exact analogue of the former problem — to determine the b_{is} $(i \geqq 2, s \geqq 2)$ so that $\Delta_{11} \neq 0$ and each $B'_{jk} = 0$ $(2 \leqq j < k)$. After suitable repetitions of the argument, the final problem is to determine the b_{is} $(i, s = n-1, n)$ so that $|b_{is}| \neq 0$ and

$$B^*_{n-1\,n} \equiv \sum_{i=n-1}^{n} b_{i\,n-1} \left\{ R^*_i b_{in} + \sum_{t \neq i}^{t=n-1, n} P^*_{it} b_{tn} \right\} = 0, \qquad P^*_{n-1\,n} = P^*_{n\,n-1} \equiv P.$$

If R^*_{n-1}, R^*_n and P all vanish, we may choose the b's to be any elements of determinant not zero. In the contrary case, we take

$$b_{n-1\,n-1} = R^*_n b_{nn} + P_{n-1\,n}, \qquad b_{n\,n-1} = -R^*_{n-1} b_{n-1\,n} - P b_{nn}.$$

Then $B^*_{n-1\,n} \equiv 0$, $|b| = R^*_n b_{nn}^2 + 2P b_{nn} b_{n-1\,n} + R^*_{n-1} b_{n-1\,n}^2 \neq 0$.

THEOREM. *If F is any field not † having a modulus $\leqq n-3$, there exists an n-ary linear transformation (b_{ij}) in F, with preassigned values of $b_{11}, b_{21}, \cdots, b_{n1}$ making $\sum_{i=1}^{n} a_i b_{i1}^2 \neq 0$, which replaces a given quadratic form $\sum_{i=1}^{n} a_i x_i^2$ by one of the type $\sum_{i=1}^{n} A_i x_i^2$ with $A_1 = \sum_{i=1}^{n} a_i b_{i1}^2$.*

* This condition is necessary for the solvability of the $C_{it} = 0$. If F has a modulus which divides $l-3$, the above determinant vanishes. Call $M_2, M_3, \cdots, M_{l-1}$ the minors of the elements $0, P_{32}, \cdots, P_{l-12}$. Then must $b_{il}(P_{2l} M_2 - P_{3l} M_3 + \cdots) = 0$. The determinant whose expansion is the second factor is seen as above to have the value 1.

† It is unnecessary for our applications to inquire whether or not this restriction on F is necessary for the validity of the theorem.

3. If F is a finite field the equation $a_1 b_{11}^2 + a_2 b_{21}^2 = 1$ has solutions b_{11}, b_{21} in F. Hence, applying the theorem of § 2 for $n = 2$, we see that any form $\sum_{i=1}^{m} a_i x_i^2$ in the $GF[p^k]$, $p > 2$, can be reduced by a succession of binary transformations to $\sum_{i=1}^{m-1} x_i^2 + a x_m^2$. Since one-half of the marks, not zero, of the $GF[p^k]$, $p > 2$, are squares, while the ratio of any two not-squares is a square, we can make $a = 1$ or a particular not-square ν.

THEOREM.* *In the $GF[p^k]$, $p > 2$, any m-ary quadratic form of non-vanishing determinant can be reduced to $\sum_{i=1}^{m} x_i^2$ or else to $\sum_{i=1}^{m-1} x_i^2 + \nu x_m^2$, ν being a particular not-square.*

4. Let F be the field R of all rational numbers. If $a_1 = \alpha_1/d$, $a_1 x_1^2 = \alpha_1 d y_1^2$, where $y_1 = x_1/d$. Hence we may assume that each a_i in (1) is an integer. If $a_1, \cdots, a_4$ are not all negative, the equation

$$\sum_{i=1}^{4} a_i b_i^2 = 1 \tag{2}$$

can be satisfied by rational values of $b_1, \cdots, b_4$. Indeed, there exist† integers $\beta_1, \cdots, \beta_4, \sigma$, not all zero, for which $\sum_{i=1}^{4} a_i \beta_i^2 - \sigma^2 = 0$. But if $\sigma = 0$ and say $\beta_1 \neq 0$, $\beta_2 \neq 0$, (2) is satisfied by

$$b_1 = \frac{\beta_1}{2\beta_2}\left(1 - \frac{1}{a_2}\right), \quad b_2 = \tfrac{1}{2}\left(1 + \frac{1}{a_2}\right), \quad b_3 = \frac{b_1}{\beta_1}\beta_3, \quad b_4 = \frac{b_1}{\beta_1}\beta_4.$$

It now follows that, if $a_1, \cdots, a_4$ are all negative, there exist rational solutions b_i of $\sum a_i b_i^2 = -1$. Hence, by the theorem of § 2, $\sum_{i=1}^{4} a_i x_i^2$ can be transformed by a quaternary substitution with rational coefficients into

$$\pm x_1^2 + \sum_{i=2}^{4} a_i' x_i^2,$$

the sign being $+$ unless $a_1, \cdots, a_4$ are all negative. We thus obtain the

* The reduction of quadratic forms in the $GF[p^k]$ was effected by the writer in a memoir on the linear groups defined by a quadratic invariant, American Journal of Mathematics, vol. 21 (1899), pp. 194, 222. (Cf. *Linear Groups*, pp. 158, 197.) For $p = 2$, there exist m-ary quadratic forms of non-vanishing discriminant D only for $m = 2n$, and then the two canonical types are

$$F_\lambda = \lambda x_1^2 + \lambda y_1^2 + \sum_{i=1}^{n} x_i y_i,$$

where $\lambda = 0$ or a particular one of the marks for which $\lambda x_1^2 + \lambda y_1^2 + x_1 y_1$ is irreducible in the $GF[2^k]$. For $k = 1$, JORDAN states in his memoir cited above that a quadratic form is reducible to F_0 or F_1 according as $\left(\frac{2}{D}\right) = +1$ or -1. This is clearly an oversight since D is the mark 1 and thus can be taken to be an arbitrary odd integer. The same oversight occurs in DE SÉGUIER'S *Groupes Abstraits*, p. 51, footnote.

† A. MEYER, Vierteljahrschrift der naturforschenden Gesellschaft in Zurich, vol. 29 (1884), pp. 209–222. He shows that $a_1 x_1^2 + \cdots + a_5 x_5^2 = 0$ has integral solutions x_i not all zero if $a_1, \cdots, a_5$ are integers $\neq 0$, not all of like sign.

THEOREM. *Within the field of all rational numbers any n-ary quadratic form of non-vanishing determinant can be transformed into one of the forms*

$$-\sum_{i=1}^{s} x_i^2 + \sum_{i=s+1}^{n-3} x_i^2 + ax_{n-2}^2 + bx_{n-1}^2 + cx_n^2,$$

where $s = 0$ *unless* a, b, c *are all negative.*

The difficult problem of ultimate canonical forms is not undertaken here. I pass to hermitian forms and effect a complete reduction to canonical forms.

II. *On the reduction of hermitian forms.*

5. Let F be any field* for which there is an equation

$$x^2 + ux + v = 0 \qquad (\text{roots } \omega, \bar{\omega}, \bar{\omega} \neq \omega)\dagger,$$

belonging to and irreducible in F. Denote the field $F(\omega)$ by Q. Any element of Q may be given the form $e = a + b\omega$, a and b in F. Set $\bar{e} = a + b\bar{\omega}$. Then

$$H_a \equiv \sum_{i,j}^{1,\cdots,n} \alpha_{ij}\xi_i\bar{\xi}_j \qquad (\text{each } \alpha_{ij} \text{ in } Q,\ \bar{\alpha}_{ij} = \alpha_{ji})$$

will be called a *hermitian form in the field* Q, and $|\alpha_{ij}|$ will be called the *determinant* of H_a. Under a linear transformation

$$\xi_i = \sum_{k=1}^{n} \beta_{ik}\eta_k, \qquad \bar{\xi}_i = \sum_{k=1}^{n} \bar{\beta}_{ik}\bar{\eta}_k \qquad (i = 1, \cdots, n), \tag{3}$$

in which the β_{ik} are elements of Q and $B \equiv |\beta_{ik}| \neq 0$, H_a becomes a hermitian form H_γ whose determinant $|\gamma_{ij}|$ equals $B\bar{B}|\alpha_{ij}|$.‡

6. THEOREM. *By a transformation* (3) *in* Q *of determinant unity, any hermitian form* H_a *in* Q *of non-vanishing determinant* Δ *can be reduced to*

$$\sum_{i=1}^{n} \gamma_i\eta_i\bar{\eta}_i \qquad (\text{each } \gamma_i \text{ in } F). \tag{4}$$

We first reduce H_a to a form $H_{a'}$ having $\alpha'_{11} \neq 0$. If $\alpha_{11} = 0$ and $\alpha_{jj} \neq 0$, we apply the transformation $\xi_1 = \eta_j$, $\xi_j = -\eta_1$. If every $\alpha_{ii} = 0$, we may take $\alpha_{12} \neq 0$; under the transformation $\xi_1 = \eta_1$, $\xi_2 = \eta_2 + \mu\eta_1$, H_a becomes $H_{a'}$ with

* In particular, F shall not be the field of all complex numbers $x + yi$, nor a field $F_{m,p}$ defined as the aggregate of the Galois fields of orders p^m, p^{2m}, p^{4m}, p^{8m}, $\cdots$. If in a field F of modulus 2 every equation $x^2 + ux + v = 0$ $(u \neq 0)$ is reducible, F contains $F_{1,2}$.

† In the field of all rational functions of a variable z with integral coefficients taken modulo 2, the equation $x^2 = z$ is irreducible but has equal roots. We exclude such a case $\bar{\omega} = \omega$, since the problem is then that of bilinear forms in F subject to cogredient transformations.

‡ The simplest proof follows by use of generators of the types

$$\xi_1 = \eta_1 + \lambda\eta_2,\ \xi_i = \eta_i \qquad \text{and} \qquad \xi_1 = B\eta_1,\ \xi_i = \eta_i \qquad (i = 2, \cdots, n).$$

$\alpha'_{11} = \alpha_{12}\bar{\mu} + \bar{\alpha}_{12}\mu$. Take $\mu = (a + b\omega)/\bar{\alpha}_{12}$; then $\alpha'_{11} \equiv 2a - bu$ can be made different from zero, since we have excluded the case $a = 0$ when F has modulus 2 by assuming that $\bar{\omega} \neq \omega$.

In $H_{\alpha'}$ with $\alpha'_{11} \neq 0$, set $\xi_1 = \eta_1 - \bar{\alpha}'_{12}\alpha'^{-1}_{11}\eta_2$, $\xi_2 = \eta_2$; there results $H_{\alpha''}$ with $\alpha''_{11} = \alpha'_{11}$, $\alpha''_{12} = \alpha''_{21} = 0$. Similarly, we make every α_{1i} and α_{i1} zero $(i > 1)$ and reach $H_{\alpha''} = \alpha'_{11}\eta_1\bar{\eta}_1 + f$, where f is a hermitian form on η_i, $\bar{\eta}_i$ $(i = 2, \cdots, n)$ of determinant $\neq 0$.

7. Let F be the $GF[p^m]$, so that Q is the $GF[p^{2m}]$. For any mark g of F the equation $c\bar{c} = g$, viz., $c^{p^m+1} = g$, is solvable in Q. *Hence there is an unique canonical form* $\Sigma\xi_i\bar{\xi}_i$.

8. Let F be the field of all real numbers and set $\omega = \sqrt{-1}$. Then $\gamma_i = \pm c_i^2$, so that (4) can be reduced to one of the forms

$$h_r \equiv \sum_{i=1}^{r} \xi_i\bar{\xi}_i - \sum_{i=r+1}^{n} \xi_i\bar{\xi}_i \qquad (r = 0, 1, \cdots, n). \tag{5}$$

Now r is an invariant, i. e., h_r can not be reduced to h_ρ $(\rho \neq r)$ by a transformation (3). Indeed, for $\xi_i = x_i + y_i\sqrt{-1}$, (3) becomes a special $2n$-ary real linear transformation, and (5) becomes

$$\sum_{i=1}^{r}(x_i^2 + y_i^2) - \sum_{i=r+1}^{n}(x_i^2 + y_i^2), \tag{6}$$

so that $2r$ is an invariant by § 1. *There are* $n + 1$ *canonical forms* (5).

9. Consider for a general field F the possible normalizations of a binary hermitian form $h = a(x\bar{x} + ry\bar{y})$, a and r in F, and each $\neq 0$. Set

$$x = \lambda X + \mu Y, \qquad y = \rho X + \sigma Y, \qquad D \equiv \lambda\sigma - \mu\rho \neq 0.$$

Then h becomes

$$h' = a(\lambda\bar{\lambda} + r\rho\bar{\rho})X\bar{X} + a(\mu\bar{\mu} + r\sigma\bar{\sigma})Y\bar{Y} + AX\bar{Y} + \bar{A}\bar{X}Y,$$

where $A = a(\lambda\bar{\mu} + r\rho\bar{\sigma})$. We desire that $A = 0$. For $\lambda \neq 0$, we take $\mu = -r\bar{\rho}\sigma/\bar{\lambda}$. Then $D = (\lambda\bar{\lambda} + r\rho\bar{\rho})\sigma/\bar{\lambda}$, and

$$h' = a(\lambda\bar{\lambda} + r\rho\bar{\rho})(X\bar{X} + r\tau\bar{\tau}Y\bar{Y}), \qquad \tau \equiv \frac{\sigma}{\lambda}.$$

The same form with $\tau = \mu/r\rho$ results if $\lambda = 0$, whence $\sigma = 0$. Hence r can be changed only by a factor $\tau\bar{\tau}$. By a preliminary unary transformation on y, we can restrict r to the series of multipliers $1, m_1, m_2, \cdots$ in a rectangular array of the elements of F with the various distinct elements $1, \kappa_1\bar{\kappa}_1, \kappa_2\bar{\kappa}_2, \cdots$ in the first row, where the κ's are arbitrary in Q. To retain this normalization, we set $\tau = 1$.

THEOREM. *Let* $1, m_1, m_2, \cdots$ *be the multipliers in a rectangular array of elements of* F, *the elements of the first row being the distinct elements* $\kappa\bar{\kappa}$, κ *ranging over* Q. *Any binary hermitian form can be reduced linearly to to* $a(x\bar{x} + ry\bar{y})$, *where* a *and* r *belong to the set* $1, m_1, m_2, \cdots$. *Two such reduced forms can be transformed into each other if and only if they have the same* r, *and the ratio of their* a'*s is a mark, not zero, expressible in the form* $\lambda\lambda + r\rho\bar{\rho}$.

10. Let F be the field R of all rational numbers. Then $Q = R(\epsilon)$, where $\epsilon^2 = \nu$, ν being a fixed integer $\neq 1$ having no square factor. Thus $\bar{\epsilon} = -\epsilon$. By § 9, we can transform $ax\bar{x} + by\bar{y}$ into $Lx\bar{x} + My\bar{y}$, where $L = a\lambda\bar{\lambda} + b\rho\bar{\rho}$, λ and ρ being any elements of Q for which $L \neq 0$. If $a = \alpha/d$, α and d being integers, then $ax\bar{x} = \alpha dz\bar{z}$, $z = x/d$. Hence we may assume that a and b are integers. Set $\lambda = \alpha + \beta\epsilon$, $\rho = \gamma + \delta\epsilon$, $\alpha, \beta, \gamma, \delta$ in R. Then

$$L \equiv a\lambda\bar{\lambda} + b\rho\bar{\rho} = a(\alpha^2 - \nu\beta^2) + b(\gamma^2 - \nu\delta^2). \tag{7}$$

As* in § 4, we can choose $\alpha, \beta, \gamma, \delta$ in R to make $L = +1$ if ν, a, b are not all negative, and $L = -1$ in the contrary case. Hence if ν is positive, any n-ary hermitian form is reducible to

$$\sum_{i=1}^{n-1} \xi_i\bar{\xi}_i + r\xi_n\bar{\xi}_n; \tag{8}$$

while, if ν is negative, it can be reduced to one of the types

$$f_s \equiv -\sum_{i=1}^{s} \xi_i\bar{\xi}_i + \sum_{i=s+1}^{n-1} \xi_i\bar{\xi}_i + r\xi_n\bar{\xi}_n \quad (s = 0, 1, \cdots, n-1). \tag{9}$$

If $r > 0$ and $s > 0$, we can transform $-\xi_s\bar{\xi}_s + r\xi_n\bar{\xi}_n$ into $+\xi_s\bar{\xi}_s + r'\xi_n\bar{\xi}_n$, and hence transform f_s into f_{s-1}. The reduced forms for ν negative are thus

$$f_0 \text{ with } r > 0; f_s (s = 0, 1, \cdots, n-1) \text{ with } r < 0, \tag{10}$$

falling into $n+1$ types, each characterized by the number of its negative terms. Hence by § 8, a form of one type cannot be reduced to one of a different type by a transformation (3). But by § 5, under a transformation (3) of determinant B, the determinant $(-1)^s r$ of f_s is multiplied by $B\bar{B}$. Hence $f_s(r)$ is reducible to $f_s(\rho)$ if and only if ρ/r is expressible in the form $B\bar{B}$.

THEOREM. *Any* n*-ary hermitian form in* $R(\sqrt{\nu})$ *with non-vanishing determinant can be reduced by a linear transformation in* $R(\sqrt{\nu})$ *to one and but one of the canonical forms:*

* The case $\sigma = 0$ may now be treated more naturally. There are then solutions $\lambda_1 \neq 0$, $\rho_1 \neq 0$ in $R(\epsilon)$ of $a\lambda_1\bar{\lambda}_1 + b\rho_1\bar{\rho}_1 = 0$. Let $\tau = \rho\lambda_1/\rho_1$. Then $L = 1$ becomes $\lambda\bar{\lambda} - \tau\bar{\tau} = a^{-1}$, and is satisfied by $\lambda = \frac{1}{2}(a^{-1} + 1)$, $\tau = \frac{1}{2}(a^{-1} - 1)$.

(I) *for* $\nu > 0$, $\sum_{i=1}^{n-1} \xi_i \bar{\xi}_i + r\xi_n \bar{\xi}_n$, *where* r *ranges over the multipliers (integers)* $1, m_1, m_2, \cdots$, *in a rectangular array of all rational numbers with those representable by* $x^2 - \nu y^2$ *in the first row;*

(II) *for* $\nu < 0$, $\sum_{i=1}^{n-1} \xi_i \bar{\xi}_i + \rho\xi_n \bar{\xi}_n$, $-\sum_{i=1}^{s} \xi_i \bar{\xi}_i + \sum_{i=s+1}^{n-1} \xi_i \bar{\xi}_i - \rho\xi_n \bar{\xi}_n$, *where* $s = 0, 1, \cdots, n-1$, *and* ρ *ranges over the multipliers (positive integers) in a rectangular array of all positive rational numbers with those representable by* $x^2 - \nu y^2$ *in the first row.*

In the examples,* (k) denotes all primes of the form k; $q_1, q_2, q_3, \cdots$, denote distinct primes; $p_1, p_2, p_3, \cdots$, denote distinct primes.

For $\nu = 2$, $r = 1, q_1, q_1q_2, q_1q_2q_3, \cdots, q_i$ ranging over $(8m+3), (8m+5)$.

For $\nu = 3$, $r = \pm 1, \pm q_1, \pm q_1q_2, \cdots, q_i$ ranging over $(12m+5), (12m+7)$.

For $\nu = 5$, $r = 1, q_1, q_1q_2, \cdots, q_i$ ranging over $(20m+3), (20m+7), (20m+13), (20m+17)$.

For $\nu < 0$, $\rho = 1, q_1, q_1q_2, q_1q_2q_3, \cdots$, where for $\nu = -1$, $q_i = (4m+3)$; for $\nu = -2$, $q_i = (8m+5), (8m+7)$; for $\nu = -3$, $q_i = (3m+2)$; for $\nu = -5$, the only limitations on the primes q_i are that no one is 5, $20m+1$, or $20m+9$, while at most one is chosen from the set 2, $(20m+3), (20m+7)$; for $\nu = -6$, no q_i is $24m+1$ or $24m+7$, while at most one is chosen from the set 2, 3, $(24m+5), (24m+11)$; for $\nu = -7$, $q_i = (28m+t)$, $t = 3, 5, 13, 17, 19, 27$.

III. *The bilinear forms invariant under a given substitution* S.

11. Let F be an arbitrarily given field. We seek all bilinear functions $\Phi = \sum_{ij} \gamma_{ij} \xi_i \eta_j$ with discriminant $D \equiv |\gamma_{ij}| \neq 0$ and coefficients in F, such that Φ is invariant under a given substitution, cogredient in the two sets of variables,

$$S: \qquad \xi_i' = \sum_{j=1}^{n} \alpha_{ij} \xi_j, \qquad \eta_i' = \sum_{j=1}^{n} \alpha_{ij} \eta_j \qquad (i = 1, \cdots, n),$$

with coefficients also in F. In the canonical form of S (with the initial variables ξ_i), the new variables fall into as many classes as there are distinct roots of the characteristic equation $\Delta(\rho) = 0$ of S. Each class is composed of one or more series, the variables of any series being transformed by S into linear functions of themselves, as follows:

$$|x_0, x_1, \cdots, x_t \qquad \rho x_0, \rho(x_1 + x_0), \cdots, \rho(x_t + x_{t-1})|.$$

* The integers representable by $x^2 - 2y^2$ are, aside from square factors, $\pm 1, \pm 2, (8n+1), (8n+7)$, and their products two, three, four, $\cdots$, at a time. For $x^2 - 3y^2$, they are $-2, -3, (12m+1), -(12m+11)$. For $x^2 + 5y^2$, they are 5, $(20m+1), (20m+9), 2p_1, p_1p_2$, and the products of these expressions two, three, $\cdots$, at a time, where $p_i = (20m+3), (20m+7)$. For $x^2 + 6y^2$, they are $(24m+1), (24m+7), p_1p_2$, and their products, where $p_i = 2, 3, 24m+5$ or $24m+11$.

Similarly, S with the initial variables η_i affects the general series of canonical variables as follows:

$$|\, y_0, y_1, \cdots, y_\tau \qquad \rho_1 y_0, \rho_1(y_1 + y_0), \cdots, \rho_1(y_\tau + y_{\tau-1})\,|.$$

In the new variables, Φ becomes a sum of functions each separately invariant, the general one, f, being bilinear in x_i, $y_j\,(i = 0, \cdots, t\,;\, j = 0, \cdots, \tau)$. Let ax_iy_j be any term of the latter, $i + j$ its *rank*, i being the rank of x_i and j the rank of y_j. Since the increment obtained from any term is of lower rank, the set of terms of maximum rank in the transform of f by S is derived from the set of terms of maximum rank in f by multiplication by $\rho\rho_1$. Unless f is identically zero, $\rho\rho_1 = 1$. But if none of the variables x appeared in the new form of Φ, its discriminant would vanish. Hence there is at least one series of variables y which S_η multiplies by ρ^{-1}. With the class C_x of all the variables which S_ξ multiplies by ρ is associated a class C_y of the variables which S_η multiplies by ρ^{-1}. Then $\Phi = [C_xC_y] + \Psi$, where $[C_xC_y]$ is bilinear in the variables of C_x, C_y, while Ψ does not contain those variables. The discriminant of $[C_xC_y]$ is a factor of D and hence is not zero; this requires that the classes C_x and C_y shall contain the same number of variables. Thus ρ^{-1} must be a root of $\Delta(\rho) = 0$ of the same multiplicity as the root ρ. A first necessary condition for an invariant Φ under S is:

(11) *The characteristic equation of S must be a reciprocal equation.*

A second condition for the existence of $[C_xC_y]$ is (JORDAN, § 13):

(12) *Classes C_x, C_y must be of like type as to number and length of series.*

12. When these conditions on S are satisfied, invariants $[C_xC_y]$ of non-vanishing discriminant exist (JORDAN, § 12, case I alone occurs for bilinear functions), the general one being $\sum_{\alpha,\beta,\ldots,\nu} a_r^{\alpha\beta} f_{\alpha\beta\ldots\nu}$, where f is a definite bilinear function and the a's are any polynomials in $\rho, \rho_1, \cdots, \rho_{\nu-1}$, for which the discriminant of $[C_xC_y]$ is not zero and the following "reality" conditions hold: Since Φ is to be equal to a function of the initial variables with coefficients in F, we must have $\Phi = \sum + \Phi_1$, where

$$\sum = [C_xC_y] + [C'_xC'_y] + \cdots + [C_x^{\nu-1}C_y^{\nu-1}],$$

$[C'_xC'_y], \cdots$, being derived from $[C_xC_y]$ by interchanging ρ and $\rho_1, \cdots$, ρ and $\rho_{\nu-1}$, respectively, $\rho_1, \cdots, \rho_{\nu-1}$, being the remaining roots of the same irreducible factor of $\Delta(\rho)$. Thus $a_r^{\alpha\beta}$ must equal a polynomial in ρ. Let these conditions be satisfied. Then (JORDAN,* § 14–16) by a linear transformation leaving the canonical form of S unaltered (not necessarily the same transformation on the y's as on the x's), we can reduce Φ to a unique canonical form.

* An obvious correction (not altering the argument) is to be made on p. 237, l. 11. The expression in brackets should be

$$y'_0(x'_\omega + C^1_{\omega-1}x'_{\omega-1} + \cdots + x'_1) + y''_0(x''_\omega + C^1_{\omega-1}x''_{\omega-1} + \cdots + x''_1) + \cdots.$$

13. Conditions (11) and (12) depend upon quantities irrational in general with respect to the initial field F. It seems desirable to proceed further and exhibit purely rational conditions for the solvability of the problem. We obtain* the

THEOREM. *The necessary and sufficient conditions that a given substitution* (α_{ij}) *with coefficients in* F *shall leave invariant one or more bilinear forms with coefficients in* F *are:* (i) *the characteristic determinant* $\Delta(\rho)$ *of* S *has a decomposition into factors irreducible in* F *of the type*

$$\Delta(\rho) = S_k^{\alpha} S_l^{\beta} \cdots R_t R_t^{\sigma} \cdots R_s^{\varepsilon} R_s^{\varepsilon\sigma} \cdots,$$

where S_k, S_l, $\cdots$ *are self-reciprocal, viz.,* $S_k(\rho) \equiv \rho^k S_k(\rho^{-1})$, *while* R_t *and* R_t^{σ} *are reciprocal;* (ii) *the invariant-factors of the matrix* $\Delta(\rho)$ *are of the form*

$$S_k^{\alpha_1} S_l^{\beta_1} \cdots R_t^{\tau_1} R_t^{\sigma\tau_1} \cdots R_s^{\varepsilon_1} R_s^{\varepsilon\sigma_1}, \quad S_k^{\alpha_2} S_l^{\beta_2} \cdots R_t^{\tau_2} R_t^{\sigma\tau_2} \cdots R_s^{\varepsilon_2} R_s^{\varepsilon\sigma_2}, \quad \cdots,$$

a pair of reciprocal factors occurring always to the same power.

When these conditions are satisfied there exist in F *bilinear forms* $\sum b_{ij}\xi_i\eta_j$ *invariant under* S_ξ, S_η. *All such invariant forms are reducible to a single one by a transformation on the* ξ*'s commutation with* S_ξ, *and a (possibly different) transformation on the* η*'s commutative with* S_η.

IV. *The quadratic forms invariant under a given substitution* S.

14. JORDAN'S treatment of the case of a field of order p, a prime, can be extended immediately to any finite field, and with certain essential modifications to any infinite field. Let F be any field.

Let ρ be a root of $f(x) = 0$, where $f(x)$ is a factor of degree k of $\Delta(x)$ and is irreducible in F. If ρ^{-1} is not a root of $f(x) = 0$, the question is essentially the same as that for bilinear forms, discussed above. Let next ρ^{-1} be a root of $f(x) = 0$, $\rho^{-1} \neq \rho$. Then $f(x) = 0$ and $x^k f(x^{-1}) = 0$ are equations belonging to and irreducible in F with a root ρ in common: hence all their roots are common and $f(x) = 0$ is a reciprocal equation. Since no root equals its reciprocal, $k = \text{even} = 2\nu$, and the roots may be designated $\rho \equiv \rho_0, \rho_1, \cdots, \rho_{2\nu-1}$ with $\rho_\nu = \rho_0^{-1}, \rho_{\nu+1} = \rho_1^{-1}, \cdots, \rho_{2\nu-1} = \rho_{\nu-1}^{-1}$. In the canonical form of S all the variables corresponding to a root ρ_i are said to form the class C_i. Thus an invariant quadratic form Φ must equal $[C_0 C_\nu] + \Psi$, where $[C_0 C_\nu]$ is bilinear in the variables of classes C_0 and C_ν, while Ψ does not contain them. Now $[C_0 C_\nu]$ which is itself invariant must have the form

$$[C_0 C_\nu] = \sum_{\alpha,\beta,r} a_r^{\alpha\beta} F_{\alpha\beta r}, \tag{13}$$

where $a_r^{\alpha\beta}$ are any constants satisfying the conditions later specified, while $F_{\alpha\beta r}$

*Cf. Transactions, vol. 3 (1902), pp. 290–292.

are perfectly definite bilinear forms derived from those given by JORDAN, pp. 240, 241, by making the following changes. For b_i, b_i^{ν}, c_i, c_i^{ν}, $\cdots$ write $b_i(\rho)$, $b_i(\rho^{-1})$, $c_i(\rho)$, $c_i(\rho^{-1})$, $\cdots$, respectively; for e of § 21 write $\rho - \rho^{-1}$. In § 20, take $b_0 = \frac{1}{2}$ if F does not have modulus 2, while for modulus 2 take

$$b_0 = -\frac{1}{\alpha_\nu}(\rho^\nu + \alpha_1\rho^{\nu-1} + \alpha_2\rho^{\nu-2} + \cdots + \alpha_{\nu-1}\rho),$$

where the α's refer to the irreducible reciprocal equation with root ρ:

$$(14)\quad E \equiv y^\nu + y^{-\nu} + \alpha_1(y^{\nu-1} + y^{-\nu+1}) + \cdots + \alpha_{\nu-1}(y + y^{-1}) + \alpha_\nu = 0.$$

If α_ν vanished, there would be a factor $y - y^{-1}$ modulo 2. Since the variables entering $F_{\alpha\beta r}$ are linear functions of the initial variables ξ_i whose coefficients are polynomials in ρ with coefficients in F, $F_{\alpha\beta r}$ can be expressed as a function of ρ and the ξ's with coefficients in F. In particular $F_{\alpha\beta r}$ is unaltered by any permutation of the roots $\rho_1, \rho_1^{-1}, \cdots, \rho_{\nu-1}, \rho_{\nu-1}^{-1}$. The same must be true of $[C_0C_\nu]$, which is composed of all the terms of Φ involving the variables of classes C_0 and C_ν (viz., those corresponding to the roots ρ and ρ^{-1}), since Φ is to be equal to a function of the ξ's with coefficients in F. Hence the a's in (13) must be symmetric functions of those $2\nu - 2$ roots; but the latter are the roots of $E/(y + y^{-1} - \rho - \rho^{-1}) = 0$, E being given by (14). *Hence the coefficients in* (13) *are rational functions of* ρ *with coefficients in* F. By its construction $F_{\alpha\beta r}$ becomes $F_{\beta\alpha r}$ when ρ is replaced by ρ^{-1}. Hence

$$(15)\qquad a_r^{\alpha\beta}(\rho) = a_r^{\beta\alpha}(\rho^{-1}),$$

whence $a_r^{\alpha\alpha}$ is a polynomial in $\rho + \rho^{-1}$.

Finally, since $[C_0C_\nu] + \Psi$ is to be equal to a function Φ of the ξ's with coefficients in F, Ψ must contain the bilinear forms $[C_1C_{\nu+1}], \cdots, [C_{\nu-1}C_{2\nu-1}]$ derived from $[C_0C_\nu]$ by replacing ρ by $\rho_1, \cdots, \rho_{\nu-1}$, respectively. Hence must $\Phi = Q + \Phi_1$, where Φ_1 involves no variable in the classes $C_0, \cdots, C_{2\nu-1}$, while

$$(16)\qquad Q \equiv \sum_{i=0}^{\nu-1}[C_iC_{\nu+i}] = \sum_{\alpha,\beta,r,i} a_r^{\alpha\beta}(\rho_i)F_{\alpha\beta r}^{(i)},$$

$F_{\alpha\beta r}^{(i)}$ being the same function of ρ_i and the variables of C_i, $C_{\nu+i}$ that $F_{\alpha\beta r}^{0} \equiv F_{\alpha\beta r}$ is of ρ_0 and the variables of C_0, C_ν. In (16), the a's are arbitrary rational functions of ρ_i satisfying (15) and making the descriminant of $[C_0C_\nu]$ not zero (requiring that certain determinants of the a's be $\neq 0$). The same argument is to be repeated for Φ_1 with reference to each new irreducible factor of $\Delta(\rho)$.

15. The next problem is the reduction of $[C_0C_\nu]$ to one or more normal forms by means of a transformatian of variables leaving S unaltered. JORDAN shows (§§ 23–28) that a unique normal form results when F is a field of order

p. The same is true for any finite field, but not for an arbitrary infinite field. For an arbitrary field F we can proceed with JORDAN's normalization by observing the following modifications.* The constants a, a^{p^ν}, λ, λ^{p^ν}, $\cdots$ are to be interpreted as $a(\rho)$, $a(\rho^{-1})$, $\lambda(\rho)$, $\lambda(\rho^{-1})$, $\cdots$, respectively.

To prove that the argument at the bottom of p. 243 remains valid, we have to show that, for $a_{mm}^{21} \neq 0$,

$$\lambda(\rho)a_{mm}^{21}(\rho) + \lambda(\rho^{-1})a_{mm}^{12}(\rho) + \lambda(\rho)\lambda(\rho^{-1})a_{mm}^{22}(\rho)$$

is not identically zero for every rational function λ. But if the sum vanished for $\lambda(\rho) = 1, -1$, and ρ, then would

$$\begin{vmatrix} 1 & 1 & 1 \\ -1 & -1 & 1 \\ \rho & \rho^{-1} & 1 \end{vmatrix} \equiv -2(\rho^{-1} - \rho) = 0,$$

which is impossible if F does not have modulus 2. For modulus 2, we employ $\lambda(\rho) = 1, \rho$, and $\rho + 1$, obtaining as the determinant $(\rho^{-1} + \rho)(\rho^{-1} - \rho)$.

The argument to make $a_{mm}^{11} = 1$ (top of p. 244) must now be abandoned. For the present we allow a_{mm}^{11} to remain arbitrary, but $\neq 0$. To make $a_{mm}^{12} = a_{mm}^{21} = 0$, apply the transformation which replaces x_i' and y_i' by $x_i' - a(\rho)x_i''$ and $y_i' - a(\rho^{-1})y_i''$, respectively, for $i = m, \cdots, 0$, taking $a(\rho) = a_{mm}^{21}(\rho)/a_{mm}^{11}(\rho)$. By analogous transformations we can make every $a_{mm}^{\alpha\beta} = 0\,(\alpha \neq \beta)$ and reach

$$\phi_a = A \sum_{\alpha=1}^{l} \{ a_{mm}^{\alpha\alpha} [x_0^\alpha y_m^\alpha + (-1)^m x_m^\alpha y_0^\alpha] \} \quad (\text{each } a_{mm}^{\alpha\alpha} \neq 0,\ A \neq 0). \tag{17}$$

Consequently we insert the factors $a_{mm}^{\alpha\alpha}$ in the formulæ of §§ 26–28. For λ on p. 248 we now take

$$\lambda(\rho) = -c(\rho^{-1}) \div \{ A(\rho^{-1}) a_{mm}^{\alpha\alpha}(\rho) \}.$$

For case 1°, p. 250, the condition to be satisfied is now

$$(-1)^{m'-r'} [\lambda(\rho^{-1}) + \lambda(\rho)] a_{mm}^{\alpha\alpha} + a_{r'}^{\alpha\alpha} = 0,$$

where each a is a rational function of $\rho + \rho^{-1}$. To this end we apply the

LEMMA. *If ρ is a root of an irreducible reciprocal equation* (14), *we can determine a polynomial $\lambda(\rho)$ such that $\lambda(\rho^{-1}) + \lambda(\rho) = f(\rho + \rho^{-1})$, f being any given polynomial, where the coefficients of λ and f belong to the arbitrary field F.*

In view of (14), we may set

$$f = g_1(\rho^{\nu-1} + \rho^{-\nu+1}) + \cdots + g_{\nu-1}(\rho + \rho^{-1}) + g_\nu \qquad (g\text{'s in } F).$$

* We do not consider the numerical results of §§ 24, 25, peculiar to finite fields. On p. 245, line 6, the term $c_1^{\mu\nu} d_1$ should be deleted. The group considered in § 24 occurs in the literature, Mathematische Annalen, vol. 52 (1899) p. 561, and vol. 55 (1902), p. 521, as the hyperorthogonal group in the $GF[p^{2\nu}]$.

We proceed to exhibit a solution λ of the form

$$\lambda(\rho) = c_0\rho^\nu + c_1\rho^{\nu-1} + \cdots + c_\nu \qquad (c\text{'s in } F).$$

If F does not have modulus 2 we may take

$$c_0 = 0,\ c_1 = g_1,\ \cdots,\ c_{\nu-1} = g_{\nu-1},\ c_\nu = \tfrac{1}{2}g_\nu.$$

If F has modulus 2, we apply (14) to eliminate $\rho^\nu + \rho^{-\nu}$ and get

$$\lambda(\rho^{-1}) + \lambda(\rho) = \sum_{i=1}^{\nu-1}(c_i - c_0\alpha_i)(\rho^{\nu-i} + \rho^{-\nu+i}) - c_0\alpha_\nu.$$

Hence this equals f if we set

$$c_0 = \frac{g_\nu}{\alpha_\nu}, \qquad c_i = g_i + \frac{\alpha_i g_\nu}{\alpha_\nu} \qquad (i = 1, \cdots, \nu-1).$$

For case 2°, page 250, the condition to be satisfied is now*

$$(-1)^{m'-r'-1}e[\lambda(\rho^{-1}) - \lambda(\rho)]\,a_m^{aa} + a_r^{aa} = 0,\ e \equiv \rho - \rho^{-1}.$$

Now $(-1)^{m'-r'}a_r^{aa}/a_m^{aa}$ equals a polynomial $f(\rho+\rho^{-1})$ by (15). Set $\mu(\rho) = (\rho^{-1} - \rho)\lambda(\rho)$. The resulting condition $\mu(\rho^{-1}) + \mu(\rho) = f(\rho+\rho^{-1})$ may be satisfied by the Lemma.

Cases 3°_1, 3°_2 are analogous to 2°, 1°, respectively.

Hence $[C_0 C_\nu]$ *can be reduced to the semi-normal form*

$$(18) \qquad \sum_{a=1}^{l} a_m^{a,} F_{aam} + \sum_{\beta=l+1}^{l+l'} a_{m'}^{\beta\beta} F_{\beta\beta m'} + \sum_{\gamma=l+l'+1}^{l+l'+l''} a_{m''}^{\gamma\gamma} F_{\gamma\gamma m''} + \cdots,$$

where each $a = a(\rho) = a(\rho^{-1}) \neq 0$, *while* F_{aam} *is a bilinear function with fixed coefficients of* x_i^a *and* y_i^a $(i = 0, 1, \cdots, m;\ \alpha = 1, \cdots, l)$, $F_{\beta\beta m'}$ *a bilinear function of* x_i^β *and* y_i^β $(i = 0, 1, \cdots, m';\ \beta = l+1, \cdots, l+l')$, *etc. Also* $y(\rho) = x(\rho^{-1})$, $m > m' > m'' \cdots$.

16. First let F be a finite field, the $GF[p^k]$. Then $F(\rho)$ is the $GF[p^{2\nu k}]$ and $\rho^{p^{\nu k}} = \rho^{-1}$ The a's belong to the $GF[p^{\nu k}]$. Apply the transformation (commutative with S) which multiplies each x_i^a by λ and each y_i^a by $\lambda^{p^{\nu k}}$. The new coefficients of F_{aam} is $a_m^{aa}\lambda^{p^{\nu k}+1}$ and hence can be made unity by choice of λ in the $GF[p^{2\nu k}]$. *For a finite field every* a *in* (18) *can be made unity, so that there is a unique normal form.*

17. For a general field F the question is not so simple; we shall have to

* By a misprint the wrong sign is given by JORDAN. On p. 249, 5th line from the bottom, $\lambda^{p^{\nu\nu}} y^a_{m-r}$ should read $\lambda^{p^{\nu\nu}} y^a_0$. On p. 250, 3d line from the bottom, $\lambda^{p^{\nu\nu}} + \lambda$ should read $\lambda^{p^{\nu\nu}} - \lambda$.

consider normalizations not used or needed in JORDAN's case. Now S affects* the variables entering (18) as follows:

$$(19)\quad \begin{cases} x_0^\alpha,\ x_1^\alpha,\ \cdots,\ x_m^\alpha, & \rho x_0^\alpha,\ \rho(x_1^\alpha+x_0^\alpha),\ \cdots,\ \rho(x_m^\alpha+x_{m-1}^\alpha) \\ y_0^\alpha,\ y_1^\alpha,\ \cdots,\ y_m^\alpha, & \rho^{-1}y_0^\alpha,\ \rho^{-1}(y_1^\alpha+y_0^\alpha),\ \cdots,\ \rho^{-1}(y_m^\alpha+y_{m-1}^\alpha) \\ & \qquad (\alpha = 1, \cdots, l) \\ x_0^\beta,\ x_1^\beta,\ \cdots,\ x_{m'}^\beta, & \rho x_0^\beta,\ \rho(x_1^\beta+x_0^\beta),\ \cdots,\ \rho(x_{m'}^\beta+x_{m'-1}^\beta) \\ y_0^\beta,\ y_1^\beta,\ \cdots,\ y_{m'}^\beta, & \rho^{-1}y_0^\beta,\ \rho^{-1}(y_1^\beta+y_0^\beta),\ \cdots,\ \rho^{-1}(y_{m'}^\beta+y_{m'-1}^\beta) \\ & \qquad (\beta = l+1, \cdots, l+l') \\ \text{analogously for } x_0^\gamma,\ y_0^\gamma,\ \cdots,\ x_{m''}^\gamma,\ y_{m''}^\gamma & \quad (\gamma = l+l'+1, \cdots, l+l'+l'') . \end{cases}$$

Any substitution P commutative with (19) is the product of a substitution T on the x's by the conjugate substitution $\overline{T}$ on the y's. Further,

$$T = T_m T_{m'} T_{m''} \cdots W,$$

where each factor is commutative with S, T_m denoting a substitution of the form

$$(20)\qquad x_i^\alpha \qquad \sum_{\delta=1}^{l} b_{\alpha\delta} x_i^\delta \qquad (\alpha = 1, \cdots, l;\ i = 0, 1, \cdots, m),$$

$T_{m'}$ being cogredient on $x_i^{l+1}, \cdots, x_i^{l+l'}$, for $i = 0, 1, \cdots, m'$, etc., while W is derived from substitutions, the general one of which replaces x_i^α by $x_i^\alpha + \lambda x_{i-m+r}^\beta$ ($i = m, \cdots, m-r$) and leaves fixed x_j^α ($j = m-r-1, \cdots, 0$), r being $< m$. Hence† T replaces each x_0^α by a linear function of $x_0', \cdots, x_0^l$ only, while

* Here and henceforth it is to be understood that the remaining variables $x(\rho_1), y(\rho_1), \cdots, x(\rho_{\nu-1}), y(\rho_{\nu-1})$, conjugate to $x(\rho) \equiv x$, undergo the transformation conjugate to that on x. Thus the complete substitution can be expressed as a substitution on the initial variables with coefficients in F.

† The explicit form of T is not essential to the argument; it is moreover fairly complex (*Linear Groups*, § 218). For $m = 2$, $m' = 1$, $l = 2$, $l' = 1$, S is

$$\begin{vmatrix} x_0^\alpha,\ x_1^\alpha,\ x_2^\alpha & \rho x_0^\alpha,\ \rho(x_1^\alpha + x_0^\alpha),\ \rho(x_2^\alpha + x_1^\alpha) \\ x_0''',\ x_1''' & \rho x_0''',\ \rho(x_1''' + x_0''') \end{vmatrix} \qquad (\alpha = 1, 2),$$

the explicit form of T is the following (zero coefficients not being entered):

	x_0'	x_0''	x_1'	x_1''	x_2'	x_2''	x_0'''	x_1'''
x_0'	b_{11}	b_{12}						
x_0''	b_{21}	b_{22}						
x_1'	c_1	c_2	b_{11}	b_{12}			e_1	
x_1''	c_3	c_4	b_{21}	b_{22}			e_2	
x_2'	d_1	d_2	c_1	c_2	b_{11}	b_{12}	f_1	e_1
x_2''	d_3	d_4	c_3	c_4	b_{21}	b_{22}	f_2	e_2
x_0'''	g_1	g_2					b_{33}	
x_1'''	h_1	h_2	g_1	g_2			k	b_{33}

$x'_m, \cdots, x^l_m$ appear in the functions by which T replaces x^a_i only when $i = m$. Analogous remarks on the y's hold for $\overline{T}$. Hence in the function by which $P \equiv T\overline{T}$ replaces (18), the terms involving the variables x^a_m and y^a_m of maximum rank m come only from the terms of (18) involving these variables. But the latter terms are given by ϕ_a of (17). Hence in order that P shall transform f_a given by (18) into a similar function $f_{a'}$, it is necessary that P shall transform ϕ_a into $\phi_{a'} + \psi$, where the variables of ψ are of rank $< m$. Hence the factor T_m, given by (20), of T must transform ϕ_a into $\phi_{a'}$. The necessary and sufficient conditions for this are

$$\sum_{\alpha=1}^{l} a^{\alpha\alpha}_m b_{\alpha\delta} \overline{b}_{\alpha\epsilon} = \begin{cases} 0 & \text{if } \delta \neq \epsilon, \\ a'^{\delta\delta}_m & \text{if } \delta = \epsilon. \end{cases} \tag{21}$$

We assume that these conditions on T_m are satisfied. Then T_m replaces $\sum_{\alpha=1}^{l} a^{\alpha\alpha}_m x^\alpha_i y^\alpha_j$ by

$$\sum_{\delta,\epsilon=1}^{l} \left(\sum_{\alpha=1}^{l} a^{\alpha\alpha}_m b_{\alpha\delta} \overline{b}_{\alpha\epsilon} \right) x^\delta_i y^\epsilon_j = \sum_{\delta=1}^{l} a'^{\delta\delta}_m x^\delta_i y^\delta_j .$$

Hence T_m replaces $\sum_{\alpha=1}^{l} a^{\alpha\alpha}_m F_{\alpha\alpha m}$ by $\sum_{\alpha=1}^{l} a'^{\alpha\alpha}_m F_{\alpha\alpha m}$. The same reasoning applies to $T_{m'}, T_{m''}, \cdots$. Hence * if $T\overline{T}$ replaces f_a by $f_{a'}$, the product $T_m \overline{T}_m T_{m'} \overline{T}_{m'} \cdots$ must replace f_a by $f_{a'}$.

THEOREM. *Any possible transformation of* (18) *into a similar function by means of a substitution commutative with* S *can be effected by the simple substitutions* (20) *subject to conditions* (21). *The normalization of* (18) *must take place in the individual sums independently.*

18. In the actual normalization of $\sum_a a^{\alpha\alpha}_m F_{\alpha\alpha m}$, the plan of § 9 is to apply, instead of a single l-ary substitution (20), a succession of binary substitutions [special cases of (20)]:

$$|x'_i \quad x''_i \quad b_{11}x'_i + b_{12}x''_i, \; b_{21}x'_i + b_{22}x''_i| \quad (i = 0, 1, \cdots, m), \tag{22}$$

where by (21), $a^{11}_m b_{11}\overline{b}_{12} + a^{22}_m b_{21}\overline{b}_{22} = 0$. It follows as in § 9 that $a(F_{11m} + rF_{22m})$ can be multiplied by $\lambda\overline{\lambda} + r\mu\overline{\mu}$ by means of a binary substitution commutative with S, where λ and μ are any rational functions of ρ such that $\lambda\overline{\lambda} + r\mu\overline{\mu} \neq 0$.

19. Finally, for a root $\rho = \pm 1$, the operations take place in the initial field F. We may therefore follow JORDAN's developments † (§§ 33–35), to obtain the most general invariant $[C]$ involving the variables with multiplier ± 1. But the reduction of $[C]$ to normal forms by means of substitutions commuta-

* For our normalization, we may therefore dispense with the substitutions W. It may be be noted in passing that W replaces $f_{a'}$ by $f_{a''}$ only when $f_{a'} \equiv f_{a''}$. This may be proved directly by noting that the first term in JORDAN's expression for F_{aam} appears with the same coefficient after transformation by W.

† On p. 258, l. 7, read $2n - 2$ in the second term; at the end of G_{an} read $x_{n-k-k'}$.

tive with S requires essential modifications for the generalization to an arbitrary field F. The first step (§ 37) is now impossible in general. Instead,* if F is any field not having modulus 2, we employ the preliminary normalization reducing ϕ to $\sum_{\alpha=1}^{l} a_n^{\alpha\alpha} x_n^\alpha x_n^\alpha$, each $a \neq 0$. Although a_n^{11} is not necessarily unity, the argument in §§ 38–40 holds after an evident modification, so that we reach

$$[C] = \sum_{\alpha=1}^{l} a_n^{\alpha\alpha} G_{\alpha n} + \Psi \qquad (\text{if } m = \text{even} = 2n).$$

No changes are necessary in §§ 41–43, so that there results

$$[C] = \sum_{k=1}^{l/2} f_{2k-1,\,2k,\,2n-1} + \Psi \qquad (\text{if } m = \text{odd} = 2n-1).$$

Indeed, all alternate bilinear forms of determinant $\neq 0$,

$$\sum_{i,\,k}^{1,\cdots,2m} c_{ik} x_i X_k \qquad (c_{ik} = -c_{ki},\ c\text{'s in field } F),$$

are reducible by a linear substitution in F, congredient on the x's and X's, to a unique normal form $\sum_{k=1}^{m} (x_{2k-1} X_{2k} - x_{2k} X_{2k-1})$.

Continuing similarly the reduction of Ψ, we obtain as a semi-canonical form of $[C]$ a sum of terms G and f, affecting different variables. The factors a of the G are as yet any marks not zero of F, but otherwise the coefficients in $[C]$ are all fixed constants. As shown above for hermitian forms, so here further normalization must take place in the separate sums

$$(23) \qquad \sigma_\alpha \equiv \sum_{\alpha=1}^{l} a_n^{\alpha\alpha} G_{\alpha n}, \qquad \sum_{\beta=l+1}^{l+l'} a_n^{\beta\beta} G_{\beta n'}, \cdots \qquad (n > n' > n'', \cdots),$$

the only substitution effective in normalizing σ_α being of the form

$$(24) \qquad x_i^\alpha \qquad \sum_{\delta=1}^{l} b_{\alpha\delta} x_i^\delta \qquad (\alpha = 1, \cdots, l;\ i = 0, 1, \cdots, n).$$

Now $G_{\alpha n} = x_n^\alpha x_n^\alpha + \sum c_{ij} x_i^\alpha x_j^\alpha (i \neq j)$. Further, (24) replaces $x_i x_j (i \neq j)$ by a sum of such terms. Hence if (24) replaces σ_α by $\sigma_{\alpha'}$ it must replace $\sum a_n^{\alpha\alpha} x_n^\alpha x_n^\alpha$, the only terms in σ_α with like subscripts to the two x's, by $\sum a_n'^{\alpha\alpha} x_n^\alpha x_n^\alpha$, the only terms in $\sigma_{\alpha'}$ with like subscripts to the two x's. Hence, since F does not have modulus 2, we derive the necessary conditions

$$(25) \qquad \sum_{\alpha=1}^{l} a_n^{\alpha\alpha} b_{\alpha\delta}^2 = a_n'^{\delta\delta}, \qquad \sum_{\alpha=1}^{l} a_n^{\alpha\alpha} b_{\alpha\delta} b_{\alpha\varepsilon} = 0 \quad (\delta, \varepsilon = 1, \cdots, l;\ \delta \neq \varepsilon).$$

*Some corrections make the reading of §§ 37–39 easier. In §§ 37, 38, read $[C]$ for $[aa]$. In § 38 the use of a_{2n}^{11} instead of a_n^{11} does not conform to the earlier notations. For T in § 39 read

$$T = \left| x_{2n}', \cdots, x_{2n-2k}' \quad x_{2n}' + \lambda x_{2k}', \cdots, x_{2n-2k}' + \lambda x_0' \right|.$$

But these are also sufficient conditions that (24) shall replace σ_a by $\sigma_{a'}$. Indeed, if (25) hold, (24) replaces $\sum_a a_a^{aa} x_i^a x_j^a \ (i \neq j)$ by

$$\sum_{\delta,\epsilon}^{1,\cdots,l} \left(\sum_{a=1}^{l} a_a^{aa} b_{a\delta} b_{a\epsilon} \right) x_i^\delta x_j^\epsilon = \sum_{\delta=1}^{l} a_a^{\prime\delta\delta} x_i^\delta x_j^\delta .$$

Hence for the purposes of normalization, G_{an} of σ_a may be replaced by its leading term $x_n^a x_n^a$ and the subscript i in (24) restricted to the value n. *The problem is therefore essentially the normalization of an l-ary quadratic form within the group of all l-ary substitutions in the field F.*

20. With regard to fields having modulus 2, we restrict our attention to a finite field, say the $GF[2^k]$. Then every mark is a square. The developments of JORDAN (§§ 46–53) hold for the $GF[2^k]$ after slight changes. In § 46 we must multiply $x_a'^2 + x_a''^2$ by λ, where λ is any fixed mark such that $\lambda x^2 + \lambda y^2 + xy$ is irreducible in the $GF[2^n]$ (cf. *Linear Groups*, § 199). Hence in § 47, $G_{1a} + G_{2a}$ must be multiplied by λ; similarly in the latter sections. In § 48 the case $b^a \neq 0$ can be reduced to the case $b^a = 1$ by applying the substitution which multiplies x_i^a by $(b^a)^{-\frac{1}{2}}$ for $i = 0, 1, \cdots, m$.

THE UNIVERSITY OF CHICAGO,
November 9, 1905.

DIE KINEMATISCHE ERZEUGUNG VON MINIMALFLÄCHEN*

ERSTE ABHANDLUNG

VON

PAUL STACKEL

1.

Eine krumme Fläche besitzt eine *kinematische Erzeugung*, wenn sie durch die Bewegung einer starren Curve nach einem bestimmten Gesetze entsteht (vergl. G. DARBOUX, *Leçons*, t. I, Livre 1, Chap. 9). Zu den so definierten Flächen, die man als kinematische Flächen bezeichnen könnte, gehören zum Beispiel die Drehungsflächen, die Schraubenflächen, die geradlinigen Flächen, die Gesimsflächen von Monge, die Translationsflächen. Einen besonderen Fall der Translationsflächen bilden die *Minimalflächen*, bei denen die erzeugenden Curven imaginäre Curven von verschwindendem Bogenelemente oder, wie S. LIE sagt, Minimalcurven sind. Dass Minimalflächen aber auch durch die Bewegung reeller Curven erzeugt werden können, zeigen die ältesten Beispiele solcher Flächen: die Umdrehungsfläche der Kettenlinie (L. EULER, 1744) und die gemeine geradlinige Schraubenfläche (CH. MEUSNIER, 1776). Ihnen fügte F. H. SCHERK (1831) die Minimalflächen hinzu, die zugleich Schraubenflächen sind, und er entdeckte auch eine Klasse von Minimalflächen, die sich durch Translation reeller Curven erzeugen lassen. Der Satz, dass diese Scherkschen Flächen nebst ihren Ausartungen die einzigen Minimalflächen sind, die durch Translation von Curven nicht verschwindenden Bogenelementes entstehen, rührt von S. LIE (1880) her; für genauere und ausführlichere Literaturangaben sei auf den Artikel: R. v. LILIENTHAL, *Besondere Flächen*, *Encyklopädie der mathematischen Wissenschaften*, Band III, 3, Heft 2/3, 1903 verwiesen.

Da alle Minimalflächen durch Translation von Minimalcurven entstehen, besitzen die soeben aufgeführten Minimalflächen die Eigenschaft, eine *mehrfache* kinematische Erzeugung zu gestatten, und man gelangt so zu der Frage, welchen Minimalflächen überhaupt diese Eigenschaft zukommt. Ihrer Untersuchung sollen die folgenden Abhandlungen gewidmet sein, und in dieser *ersten* Abhandlung das Problem behandelt werden, *welche Minimalflächen mehrfache*

* Presented to the Society February 24, 1906. Received for publication January 2, 1906.

Erzeugung durch Translation gestatten. Da durch einen jeden regulären Flächenpunkt nur zwei auf der Fläche liegende Minimalcurven hindurchgehen, gibt es nur zwei Arten der Erzeugung dieser Flächen durch Translation von Minimalcurven, je nachdem man nämlich die eine oder die andere Schar der auf der Fläche liegenden Minimalcurven benutzt, wobei dann die Curven der anderen Schar die Rolle der Leitcurven spielen. Weitere Erzeugungen durch Translation sind demnach nur möglich, wenn man Curven nicht verschwindenden Bogenelementes einer Translation unterwirft. Um die Curven zu erhalten, die auf diese Weise Minimalflächen liefern, kann man das allgemeinere Problem untersuchen, alle Flächen zu bestimmen, die überhaupt auf mehr als zwei Arten durch Translation erzeugt werden können. Mit diesem allgemeineren Probleme hat sich S. LIE in den Jahren von 1872 bis 1896 eingehend beschäftigt, und er hat schliesslich gezeigt, wie man mittels der zu einer beliebigen ebenen Curve vierter Ordnung gehörigen Abelschen Integrale alle Flächen jener Eigenschaft finden kann; ein Verzeichnis der betreffenden Abhandlungen und eine neue, elegante Herleitung des Schlussergebnisses findet man bei G. SCHEFFERS, Acta mathematica, Band 28, S. 65 bis 66. Wie LIE selbst hervorgehoben hat, ergeben sich aus seinem allgemeinen Theoreme die Minimalflächen, welche mehrfache Translationsflächen sind, wenn die Curve vierter Ordnung, die man sich in der unendlich fernen Ebene zu denken hat, in einen Kegelschnitt und den imaginären Kugelkreis zerfällt. Die Ausführung der betreffenden Rechnungen bildet einen Teil der Dissertation von R. KUMMER, Leipzig, 1894, der auch von der speciellen SCHERKSCHEN Fläche $e^y = \sin x : \sin z$ ein Modell angefertigt hat, das sich in dem Besitze des Mathematischen Institutes der Universität Leipzig befindet.

Obgleich demnach die Aufgabe, alle Curven zu bestimmen, durch deren Translation Minimalflächen entstehen, als gelöst angesehen werden kann, bietet sich doch Veranlassung darauf zurückzukommen, denn es erscheint der Wunsch berechtigt, dass man eine directe Lösung angebe, bei der keine fernliegenden Hilfsmittel gebraucht werden, und man darf auch hoffen, dass sich daraus Anhaltspunkte für die Untersuchung des allgemeineren Problems ergeben werden, die Minimalflächen zu bestimmen, die einer mehrfachen kinematischen Erzeugung fähig sind. Eine Methode zu einer solchen directen Lösung habe ich in einer Note angegeben, die Herr D. HILBERT im Juli 1905 der Göttinger Gesellschaft der Wissenschaften vorgelegt hat. Die ausführliche Entwickelung und einige Ergänzungen sollen im Folgenden gegeben werden.

2.

Um einen Ansatz für die Bestimmung aller Curven zu gewinnen, durch deren Translation Minimalflächen entstehen können, hat man auszudrücken, dass eine Translationsfläche Minimalfläche sein soll. Wird eine solche Fläche durch die

Translation einer Curve C_1 erzeugt, so gestattet sie auch immer die Erzeugung durch eine zweite Curve C_2, und man erhält so für die rechtwinkligen cartesischen Coordinaten eines Flächenpunktes die Darstellung:

$$(1)\qquad \begin{cases} x = \int f(u)\,du + \int k(v)\,dv, \\ y = \int g(u)\,du + \int l(v)\,dv, \\ z = \int h(u)\,du + \int m(v)\,dv. \end{cases}$$

Aus diesen Gleichungen ergeben sich für die Fundamentalgrössen erster und zweiter Ordnung, die mit GAUSS durch E, F, G; D, D', D'' bezeichnet werden mögen, die Werte:

$$(2)\qquad \begin{cases} E = f^2 + g^2 + h^2, \\ F = fk + gl + hm, \\ G = k^2 + l^2 + m^2, \end{cases}$$

und

$$(3)\qquad \begin{cases} D = (hg' - gh')k + (fh' - hf')l + (gf' - fg')m, \\ D' = 0, \\ D'' = -f(ml' - lm') - g(km' - mk') - h(lk' - kl'), \end{cases}$$

wo f', g', h' die Ableitungen von f, g, h nach u und k', l', m' die Ableitungen von k, l, m nach v bedeuten.

Wenn die Fläche (1) eine Minimalfläche ist, so verschwindet ihre mittlere Krümmung, und es ist daher

$$ED'' - 2FD' + GD = 0,$$

also wegen (3):

$$(4)\qquad ED'' + GD = 0.$$

Die Gleichung (4) ist erfüllt, wenn man

$$E = 0, \qquad G = 0$$

setzt. Dann haben die erzeugenden Curven C_1 und C_2 verschwindendes Bogenelement, und die Fläche entsteht durch Translation von irgend zwei Minimalcurven; man gelangt auf diese Art zu *allen* Minimalflächen. Verschwindet nur E, während G von Null verschieden ist, so folgt aus (4), dass $D = 0$ ist, mithin ist auch $DD'' - D'^2 = 0$, das Krümmungsmass verschwindet, und man erhält die *Ebene*. Zu demselben Ergebnisse kommt man, wenn G verschwindet, während E von Null verschieden ist. Wird also der triviale Fall der Ebene ausgeschlossen, so lassen sich diejenigen Minimalflächen, die nicht nur durch Translation von Minimalcurven, sondern auch durch Translation von Curven mit nicht verschwindendem Bogenelemente erzeugt werden können, characterisieren durch die Functionalgleichung:

$$(A)\quad \left\{ 0 = \frac{hg' - gh'}{f^2+g^2+h^2}\cdot k + \frac{fh' - hf'}{f^2+g^2+h^2}\cdot l + \frac{gf' - fg'}{f^2+g^2+h^2}\cdot m \right.$$
$$\left. - f\cdot\frac{ml' - lm'}{k^2+l^2+m^2} - g\cdot\frac{km' - mk'}{k^2+l^2+m^2} - h\cdot\frac{lk' - kl'}{k^2+l^2+m^2} \right.$$

deren vollständige Lösung zunächst entwickelt werden muss.

3.

Die Gleichung (A), aus der die Functionen f, g, h von u und k, l, m von v zu ermitteln sind, gehört zu den Functionalgleichungen der Form

$$(F)\qquad f_1(u)k_1(v) + f_2(u)k_2(v) + \cdots + f_n(u)k_n(v) = 0,$$

die in der Theorie der krummen Flächen nicht selten auftreten. Für $n = 4$ hatte ihre Lösungen J. A. SERRET [Journ. de math. (1) 8 (1843)] angegeben, und der allgemeine Fall war dann von G. VIVANTI [Rend. circ. Palermo, 6 (1892)] und F. PROBST (Dissertation, Würzburg 1893) behandelt worden. Am einfachsten gelangt man wohl zu der Lösung von (F), indem man den Begriff eines *r-gliedrigen Systems von n Functionen* $f_1(u), f_2(u), \cdots, f_n(u)$ einführt, nämlich ein solches System 0-, 1-, $\cdots$, $(n-1)$-gliedrig nennt, je nachdem zwischen den n Functionen genau $n, n-1, \cdots, 1, 0$ von einander unabhängige lineare homogene Relationen mit constanten, also die Veränderliche u nicht enthaltenden Coefficienten bestehen, und beachtet, dass sich die Functionen eines r-gliedrigen Systemes in der Gestalt darstellen lassen:

$$(P)\qquad f_\nu(u) = s_{\nu 1}p_1(u) + s_{\nu 2}p_2(u) + \cdots s_{\nu r}p_r(u) \quad (\nu = 1, 2, \cdots, n),$$

wobei erstens die r Functionen $p_1, p_2, \cdots, p_r$ ein r-gliedriges System bilden, also, wie man zu sagen pflegt, linear unabhängig sind, und zweitens von den Determinanten r-ter Ordnung, die man aus je r Zeilen der Matrix der rn Constanten $s_{11}, \cdots, s_{nr}$ bilden kann, mindestens eine von Null verschieden ist. Man erkennt leicht, dass umgekehrt jedes System von Functionen $f_1(u), \cdots, f_n(u)$, die sich in der Gestalt (P) darstellen lassen und zwar so, dass diese beiden Bedingungen erfüllt sind, genau r-gliedrig ist.

Bei der Lösung der Functionalgleichung (F) hat man $n+1$ wesentlich verschiedene Fälle zu unterscheiden, je nachdem nämlich das System der Functionen $f_1(u), \cdots, f_n(u)$ 0-, 1-, $\cdots$, $(n-1)$-, n-gliedrig ist. Ist es etwa r-gliedrig, so folgt aus (F), wenn die Darstellung (P) eingeführt wird, eine lineare homogene Relation zwischen den r linear unabhängigen Functionen $p_1, \cdots, p_r$, die nur dann erfüllt sein kann, wenn die Coefficienten dieser r Functionen alle identisch verschwinden. Auf diese Weise erhält man aus (F) die r Gleichungen:

$$s_{1\rho}k_1(v) + s_{2\rho}k_2(v) + \cdots + s_{n\rho}k_n(v) = 0 \quad (\rho = 1, 2, \cdots, r),$$

die wegen der Determinantenbedingung r von einander unabhängige lineare homogene Relationen zwischen den n Functionen $k_1(v), \cdots, k_n(v)$ sind. Hieraus ist zu schliessen, dass das System dieser n Functionen höchstens $(n-r)$-gliedrig ist. Ist es $(n-r)$-gliedrig, so lassen sich die Functionen $k_1(v), \cdots, k_n(v)$ als lineare homogene Functionen von $n-r$ linear unabhängigen Functionen $q_1(v), \cdots, q_{n-r}(v)$ darstellen, deren Coefficienten die Determinantenbedingung erfüllen. Ist die Determinantenbedingung nicht erfüllt, so bilden die n Functionen eine System, das weniger als $(n-r)$-gliedrig ist, und man beweist leicht, dass man *alle* $(n-r-1)$-, $(n-r-2)$-, $\cdots$, 1-, 0-gliedrigen Systeme von n Functionen bekommt, indem man die Coefficienten in der Darstellung durch die $q_1, \cdots, q_{n-r}$ geeignet wählt, oder, mit anderen Worten, dass die $(n-r)$-gliedrigen Systeme die weniger gliedrigen Systeme als Ausartungen in sich schliessen. Hieraus aber folgt, dass die Functionalgleichung (F), wenn die $f_1, \cdots, f_n$ ein r-gliedriges System bilden, in allgemeinster Weise erfüllt wird, indem man die $k_1, \cdots, k_n$ als beliebige lineare homogene Functionen von irgend welchen $n-r$ Functionen $q_1(v), \cdots, q_{n-r}(v)$ ansetzt, die nun auch nicht mehr linear unabhängig zu sein brauchen.

4.

Bei der Functionalgleichung (A) ist $n=6$, und man hat zu setzen:

$$f_1 = \frac{hg' - gh'}{f^2+g^2+h^2}, \qquad f_2 = \frac{fh' - hf'}{f^2+g^2+h^2}, \qquad f_3 = \frac{gf' - fg'}{f^2+g^2+h^2},$$

$$f_4 = f, \qquad f_5 = g, \qquad f_6 = h;$$

$$k_1 = k, \qquad k_2 = l, \qquad k_3 = m,$$

$$k_4 = \frac{lm' - ml'}{k^2+l^2+m^2}, \qquad k_5 = \frac{km' - mk'}{k^2+l^2+m^2}, \qquad k_6 = \frac{lk' - kl'}{k^2+l^2+m^2}.$$

Das System der Functionen $f_1, f_2, \cdots, f_6$ kann hier 0-, 1-, $\cdots$, 5-, 6-gliedrig sein, und dann ist das System der Functionen $k_1, k_2, \cdots, k_6$ höchstens 6-, 5-, $\cdots$, 1-, 0-gliedrig. Nun behält aber die Gleichung (1) ihre Form, wenn man darin u und v vertauscht, und dasselbe gilt von der Functionalgleichung (A). Mithin darf man unbeschadet der Allgemeinheit annehmen, dass das System der Functionen $f_1, f_2, \cdots, f_6$ höchstens dreigliedrig sei. Ist es 0-, 1-, 2-gliedrig, so besteht zwischen f, g, h mindestens eine lineare homogene Relation, mithin genügen die Coordinaten der erzeugenden Curve C_1 mindestens einer linearen Gleichung und diese Curve ist *eben*. Dasselbe findet aber auch statt, wenn zwar das System der $f_1, f_2, \cdots, f_6$ dreigliedrig ist, das System der f, g, h aber höchstens zweigliedrig ausfällt. Dann und nur dann, wenn das System der f, g, h dreigliedrig ist, erhält man eine eigentliche Raumcurve C_1. Es erweist sich als zweckmässig, den Fall einer ebenen Curve C_1 zuerst für sich

zu erledigen und darauf den erheblich schwierigeren Fall zu betrachten, dass C_1 eine eigentliche Raumcurve ist.

5.

Legt man die y-Achse in die Ebene der Curve C_1, so hat die Gleichung der Ebene die Form:

$$\sin\alpha\cdot x - \cos\alpha\cdot y = 0,$$

und es ist daher auch

$$\sin\alpha\cdot f(u) - \cos\alpha\cdot h(u) = 0,$$

also

$$f(u) = \cos\alpha\cdot U(u), \qquad h(u) = \sin\alpha\cdot U(u).$$

Nun kann $U(u)$ nicht identisch verschwinden, denn sonst wäre C_1 eine Gerade und die dadurch erzeugte Minimalfläche eine Ebene, mithin darf man

$$u_1 = \int U(u)\,du$$

als neuen Parameter einführen und erhält, wenn der Index 1 wieder fortgelassen wird:

$$f(u) = \cos\alpha, \qquad h(u) = \sin\alpha.$$

Werden diese Werte von f und h in die Functionalgleichung (A) eingesetzt, so ergibt sich die Gleichung

$$(A^*) \qquad 0 = \frac{g'}{1+g^2}\cdot(\sin\alpha\cdot k - \cos\alpha\cdot m) - \frac{\cos\alpha(ml' - lm')}{k^2+l^2+m^2} - g\,\frac{km' - mk'}{k^2+l^2+m^2} - \frac{\sin\alpha(lk' - kl')}{k^2+l^2+m^2}.$$

Aus (A^*) folgt durch Differentiation nach u:

$$(5) \qquad 0 = \frac{d}{du}\frac{g'}{1+g^2}\cdot(\sin\alpha\cdot k - \cos\alpha\cdot m) - g'\,\frac{km' - mk'}{k^2+l^2+m^2}.$$

Nun ist g' von Null verschieden, denn sonst wären f, g, h alle drei constant, und die Curve C_1 wäre eine Gerade. Ebenso kann $\sin\alpha\cdot k - \cos\alpha\cdot m$ nicht verschwinden, denn sonst lägen die beiden durch einen Punkt der Fläche gehenden erzeugenden Curven in einer und derselben Ebene, wenn man also die eine davon festhält und die andere daran parallel verschiebt, erhielte man als Fläche diese Ebene selbst. Mithin darf man (5) durch $g'\cdot(\sin\alpha\cdot k - \cos\alpha\cdot m)$ dividieren und gelangt so zu der Gleichung:

$$(6) \qquad \frac{1}{g'}\frac{d}{du}\frac{g'}{1+g^2} = \frac{km' - mk'}{(\sin\alpha\cdot k - \cos\alpha\cdot m)(k^2+l^2+m^2)},$$

aus der folgt, dass die Ausdrücke links und rechts sich auf eine und dieselbe *Constante* A reducieren.

Ist die Constante $A = 0$, so verschwindet $km' - mk'$, folglich besteht zwischen m und k eine lineare homogene Relation und daher zwischen den Coordi-

naten der Curve C_2 eine lineare Gleichung, das heisst aber, auch die Curve C_2 ist *eben*.

Ist die Constante A von Null verschieden, so wird nach (6), wenn B eine neue Constante bedeutet:

$$\frac{g'}{1+g^2} = Ag + B.$$

Wird dieser Ausdruck für $g' : (1+g^2)$ und der aus (6) folgende Wert

$$\frac{1}{A}\,\frac{km' - mk'}{k^2 + l^2 + m^2}$$

von $\sin\alpha \cdot k - \cos\alpha \cdot m$ in die Gleichung (A^*) eingetragen, so erhält man:

$$(7) \qquad -\cos\alpha(ml' - lm') + \frac{B}{A}(km' - mk') - \sin\alpha(lk' - kl') = 0,$$

aus dieser Gleichung aber lässt sich folgern, dass zwischen k, l, m eine lineare homogene Relation mit constanten Coefficienten besteht, sodass die Curve C_2, auch wenn A nicht verschwindet, *eben* ist.

Dass aus (7) die Existenz einer linearen homogenen Relation zwischen k, l, m folgt, beruht auf dem allgemeinen Satze:

Sind a, b, c Constanten, die nicht alle drei verschwinden, und besteht für die Functionen k, l, m von v die Identität:

$$(J) \qquad a(ml' - lm') + b(km' - mk') + c(lk' - kl') = 0,$$

so sind k, l, m durch eine lineare homogene Relation mit constanten Coefficienten verbunden.

Zum Beweise differentiiere man (J) nach v, wodurch die weitere Identität

$$(K) \qquad a(ml'' - lm'') + b(km'' - mk'') + c(lk'' - kl'') = 0$$

entsteht, und ordne (J) nach k', l', m', (K) aber nach k'', l'', m''. Die Ableitungen von k, l, m erhalten dann beziehungsweise dieselben Coefficienten $cl - bm$, $am - ck$, $bk - al$, und es sind also entweder die k', l', m' den k'', l'', m'' oder die $cl - bm$, $am - ck$, $bk - al$ den Determinanten der Matrix

$$\begin{pmatrix} k' & l' & m' \\ k'' & l'' & m'' \end{pmatrix}$$

proportional. In beiden Fällen verschwindet aber die Determinante

$$\begin{vmatrix} k & l & m \\ k' & l' & m' \\ k'' & l'' & m'' \end{vmatrix},$$

und das bedeutet bekanntlich, dass zwischen k, l, m eine lineare homogene Relation besteht.

Wenn die Curven C_1 und C_2 beide *eben* sind, so dürfen sie nicht in parallelen Ebenen liegen, da man sonst, wie soeben gezeigt wurde, eine Ebene erhalten würde. Mithin darf man die Schnittlinie der Curvenebenen zur y-Achse machen und erhält dann von vorn herein, indem wie u bei der Curve C_1 jetzt auch bei der Curve C_2 der Parameter v geeignet gewählt wird:

$$k = \cos\beta, \qquad m = \sin\beta.$$

Dann aber wird nach (6) $A = 0$. Folglich genügt g der Gleichung

$$g' : (1 + g^2) = B,$$

und es ist daher

$$g(u) = \operatorname{tg}(Bu + C),$$

wo C eine neue Constante bezeichnet, die jedoch unbeschadet der Allgemeinheit sofort gleich Null gesetzt werden darf, denn man darf statt u als Parameter $u_1 = u + C : B$ einführen, wobei f und h ihre Werte behalten, während g in $\operatorname{tg} Bu_1$ übergeht.

Um noch $l(v)$ zu bestimmen, setze man die gefundenen Werte von f, g, h; k, m in die Gleichung (A^*) ein. Dann ergibt sich nach Division mit der von Null verschiedenen Grösse $\sin(\alpha - \beta)$ die Gleichung:

$$0 = B + l' : (1 + l^2),$$

aus der man

$$l(v) = -\operatorname{tg}(Bv + D)$$

erhält, wo aber die Constante D wieder sofort gleich Null gesetzt werden darf.

Durch Ausführung der Quadraturen gewinnt man endlich für diejenigen Minimalflächen, die sich durch die Translation einer ebenen Curve von nicht verschwindendem Bogenelemente erzeugen lassen, die Darstellung:

$$x = \cos\alpha \cdot u + \cos\beta \cdot v,$$

$$y = -\frac{1}{B}\log\cos Bu + \frac{1}{B}\log\cos Bv,$$

$$z = \sin\alpha \cdot u + \sin\beta \cdot v.$$

Um sie zu vereinfachen, bedenke man, dass es genügt, von den einander ähnlichen Flächen der gesuchten Gattung je einen Repräsentanten anzugeben, sodass man x, y, z, durch Bx, By, Bz, ersetzen darf. Werden ferner statt u und v als Parameter $u : B$ und $v : B$ eingeführt und wird noch die xz-Ebene um die festgehaltene y-Achse um den Winkel $\frac{1}{2}(\alpha - \beta)$ gedreht, so ergibt sich die Darstellung

$$(M) \qquad \begin{cases} x = \sin\gamma \cdot (u + v), \\ y = -\log\cos u + \log\cos v, \\ z = \cos\gamma \cdot (u - v), \end{cases}$$

bei der γ das Complement von $\frac{1}{2}(\alpha + \beta)$ bezeichnet.

Die Gleichungen (M) zeigen, dass *die gefundenen Flächen* die *Scherkschen Minimalflächen*

$$(M^*) \qquad e^y = \frac{\cos(\rho x - rz)}{\cos(\rho x + rz)} \qquad (\rho = 1:2\sin\gamma,\ r = 1:2\cos\gamma)$$

und die durch Aehnlichkeitstransformationen daraus hervorgehenden Flächen sind; für $\gamma = \pi/4$ entsteht die besondere von E. KUMMER modellierte Fläche.

6.

Nunmehr soll angenommen werden, dass die Curve C_1 eine *eigentliche Raumcurve sei*, sodass die Functionen f, g, h, ein dreigliedriges System bilden, also linear unabhängig sind. Da die Functionen $f_1, f_2, \cdots, f_6$ auch ein dreigliedriges System bilden sollten, darf man zu ihrer Darstellung (P) als linear unabhängige Functionen p_1, p_2, p_3 gerade f, g, h selbst wählen und erhält dann f_1, f_2, f_3 als lineare homogene Functionen von f, g, h, oder:

$$(B) \qquad \begin{cases} \dfrac{hg' - gh'}{f^2 + g^2 + h^2} = s_{11}f + s_{12}g + s_{13}h, \\ \dfrac{fh' - hf'}{f^2 + g^2 + h^2} = s_{21}f + s_{22}g + s_{23}h, \\ \dfrac{gf' - fg'}{f^2 + g^2 + h^2} = s_{31}f + s_{32}g + s_{33}h. \end{cases}$$

Werden diese Ausdrücke in (A) eingesetzt, so folgt, weil f, g, h linear unabhängig sind, dass k, l, m den Gleichungen genügen müssen:

$$(C) \qquad \begin{cases} \dfrac{ml' - lm'}{k^2 + l^2 + m^2} = s_{11}k + s_{21}l + s_{31}m, \\ \dfrac{km' - mk'}{k^2 + l^2 + m^2} = s_{12}k + s_{22}l + s_{32}m, \\ \dfrac{lk' - kl'}{k^2 + l^2 + m^2} = s_{13}k + s_{23}l + s_{33}m. \end{cases}$$

Sind umgekehrt die sechs Gleichungen (B) und (C) erfüllt, so besteht auch die Gleichung (A). Da (B) und (C) dieselbe Structur haben, wird die Untersuchung von (B), die jetzt vorgenommen werden soll, alles liefern, was zur Lösung von (A) erforderlich ist.

Die noch unbekannten Werte der neun Constanten $s_{11}, \cdots, s_{33}$ sind jedenfalls so beschaffen, dass die Determinante

$$S = \begin{vmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{vmatrix}$$

von Null verschieden ist. Wäre nämlich $S=0$, so gäbe es drei Constanten a, b, c, die nicht alle verschwinden, sodass identisch

$$a(gh'-hg')+b(hf''-fh')+c(fg'-gf'')=0$$

ist, zwischen f, g, h bestände also eine lineare homogene Relation und die Curve C_1 wäre, gegen die Voraussetzung, eben. Auf der anderen Seite behalten aber die Gleichungen (A) ihren Sinn, auch wenn die Determinante S verschwindet. Sie liefern dann Minimalflächen, die durch Translation einer *ebenen* Curve nicht verschwindenden Bogenelementes entstehen und sich, da man bei stetiger Variation der Constanten $s_{11}, \cdots, s_{33}$ von einem beliebigen Werte der Determinante S stetig zu dem Werte Null übergehen kann, den Minimalflächen einreihen lassen, die bei einem von Null verschiedenen S entstehen; es wird sich später zeigen, dass man auf diese Weise genau zu den Flächen gelangt, die bereits gefunden worden sind, nämlich zu den Scherkschen Minimalflächen, dass sich aber auch, wenn S von Null verschieden ist, aus den Gleichungen (B) und (C) wiederum die Scherkschen Minimalflächen ergeben, ein Paradoxon, das sich dadurch aufklärt, dass diese Flächen auf unendlich viele Arten durch Translation von Curven erzeugt werden können und dass sich, abgesehen von den gemeinen geradlinigen Schraubenflächen, die zwar Ausartungen der Scherkschen Flächen sind, aber einen Platz für sich einnehmen, unter den unendlich vielen erzeugenden Curven auch immer ebene Curven befinden.

7.

Wenn man die Gleichungen (B) der Reihe nach mit f, g, h multipliciert und addiert, so ergibt sich die von den Ableitungen freie Gleichung:

$$(D)\quad s_{11}f^2+s_{22}g^2+s_{33}h^2+(s_{12}+s_{21})fg+(s_{23}+s_{32})gh+(s_{31}+s_{13})hf=0;$$

sie ist, nebenbei bemerkt, die Gleichung in homogenen Coordinaten des Kegelschnittes, der bei S. Lie den einen Bestandteil der zerfallenden Curve vierter Ordnung bildet. Die Gleichung (D) kann nicht identisch in f, g, h bestehen, denn dann wäre

$$s_{11}=0,\ s_{22}=0,\ s_{33}=0,\ s_{21}=-s_{12},\ s_{32}=-s_{23},\ s_{31}=-s_{13}$$

und daher $S=0$. Sie lässt sich erheblich vereinfachen, wenn das Coordinatensystem der x, y, z geeignet gewählt wird.

Da eine Translation der Achsen keinen Einfluss auf die f, g, h hat, soll eine Drehung vorgenommen, also die orthogonale Substitution

$$x=\alpha_1\xi+\alpha_2\eta+\alpha_3\zeta,$$
$$y=\beta_1\xi+\beta_2\eta+\beta_3\zeta,$$
$$z=\gamma_1\xi+\gamma_2\eta+\gamma_3\zeta$$

gemacht werden. Wenn die Coordinaten der Curve C_1 in dem Systeme der x, y, z die Werte hatten:

$$x_1 = \int f(u)du, \qquad y_1 = \int g(u)du, \qquad z_1 = \int h(u)du,$$

so ergibt sich jetzt für die Coordinaten ξ_1, η_1, ζ_1 die Darstellung

$$\xi_1 = \int \phi(u)du, \qquad \eta_1 = \int \chi(u)du, \qquad \zeta_1 = \int \psi(u)du,$$

in der

$$\phi(u) = \alpha_1 f(u) + \beta_1 g(u) + \gamma_1 h(u),$$
$$\chi(u) = \alpha_2 f(u) + \beta_2 g(u) + \gamma_2 h(u),$$
$$\psi(u) = \alpha_3 f(u) + \beta_3 g(u) + \gamma_3 h(u)$$

zu setzen ist. Vermöge der Gleichungen (B) erhält man für die Functionen ϕ, χ, ψ die äquivalenten Gleichungen:

$$\frac{\psi\chi' - \chi\psi'}{\phi^2 + \chi^2 + \psi^2} = \sigma_{11}\phi + \sigma_{12}\chi + \sigma_{13}\psi,$$
$$\frac{\phi\psi' - \psi\phi'}{\phi^2 + \chi^2 + \psi^2} = \sigma_{21}\phi + \sigma_{22}\chi + \sigma_{23}\psi,$$
$$\frac{\chi\phi' - \phi\chi'}{\phi^2 + \chi^2 + \psi^2} = \sigma_{31}\phi + \sigma_{32}\chi + \sigma_{33}\psi.$$

in denen die Coefficienten $\sigma_{11}, \cdots, \sigma_{33}$ folgende Werte haben:

$$\sigma_{ij} = (s_{11}\alpha_j + s_{12}\beta_j + s_{13}\gamma_j)\alpha_i + (s_{21}\alpha_j + s_{22}\beta_j + s_{23}\gamma_j)\beta_i + (s_{31}\alpha_j + s_{32}\beta_j + s_{33}\gamma_j)\gamma_i \quad (i, j = 1, 2, 3).$$

Setzt man jetzt

$$\sigma_{12} = 0, \qquad \sigma_{13} = 0,$$

so ergeben sich für die Richtungscosinus $\alpha_1, \beta_1, \gamma_1$ der ξ-Achse zwei lineare homogene Gleichungen, aus denen man in der bekannten Weise die Verhältnisse der $\alpha_1, \beta_1, \gamma_1$ berechnen kann, vorausgesetzt, dass nicht alle Determinanten zweiter Ordnung der Matrix

$$\begin{pmatrix} s_{11}\alpha_2 + s_{12}\beta_2 + s_{13}\gamma_2, & s_{21}\alpha_2 + s_{22}\beta_2 + s_{23}\gamma_2, & s_{31}\alpha_2 + s_{32}\beta_2 + s_{33}\gamma_2 \\ s_{11}\alpha_3 + s_{12}\beta_3 + s_{13}\gamma_3, & s_{21}\alpha_3 + s_{22}\beta_3 + s_{23}\gamma_3, & s_{31}\alpha_3 + s_{32}\beta_3 + s_{33}\gamma_3 \end{pmatrix}$$

identisch verschwinden. Für diese Determinanten ergeben sich aber, wenn die adjungierten Grössen von $s_{11}, \cdots, s_{33}$ mit $t_{11}, \cdots, t_{33}$ bezeichnet und die Relationen

$$\beta_2\gamma_3 - \gamma_2\beta_3 = \alpha_1, \qquad \gamma_2\alpha_3 - \alpha_2\gamma_3 = \beta_1, \qquad \alpha_2\beta_3 - \alpha_3\beta_2 = \gamma_1,$$

benutzt werden, die Werte

$$t_{11}\alpha_1 + t_{12}\beta_1 + t_{13}\gamma_1, \qquad t_{21}\alpha_1 + t_{22}\beta_1 + t_{23}\gamma_1, \qquad t_{31}\alpha_1 + t_{32}\beta_1 + t_{33}\gamma_1,$$

und diese können, da ja die Richtungscosinus α_1, β_1, γ_1 sicher nicht alle drei gleich Null sind, nur dann sämtlich verschwinden, wenn die Determinante der $t_{11}, \cdots, t_{33}$ verschwindet; diese ist aber gleich S^2. Mithin folgt aus den Gleichungen

$$\sigma_{12} = 0, \qquad \sigma_{13} = 0,$$

dass

$$\alpha_1 = \mu(t_{11}\alpha_1 + t_{12}\beta_1 + t_{13}\gamma_1),$$
$$\beta_1 = \mu(t_{21}\alpha_1 + t_{22}\beta_1 + t_{23}\gamma_1),$$
$$\gamma_1 = \mu(t_{31}\alpha_1 + t_{32}\beta_1 + t_{33}\gamma_1)$$

ist, wo man den Proportionalitätsfactor μ aus der Gleichung

$$\begin{vmatrix} \mu t_{11} - 1 & \mu t_{12} & \mu t_{13} \\ \mu t_{21} & \mu t_{22} - 1 & \mu t_{23} \\ \mu t_{31} & \mu t_{32} & \mu t_{33} - 1 \end{vmatrix} = 0,$$

zu bestimmen hat. Hieraus ergeben sich aber, da das absolute Glied gleich -1 ist, stets von Null verschiedene Werte des Factors μ. Da ferner der Coefficient von μ^3 gleich S^2, also von Null verschieden ist, so gibt es bei reellen Werten der $s_{11}, \cdots, s_{33}$ immer mindestens einen reellen Wert von μ, also auch mindestens ein System reeller Werte der α_1, β_1, γ_1 und daher auch mindestens eine reelle Transformation der Coordinaten x, y, z, bei der $\sigma_{12} = 0$, $\sigma_{13} = 0$ wird.

Hiermit ist bewiesen, dass man von vornherein annehmen darf, die Constanten s_{12} and s_{13} seien gleich Null.

Man könnte jetzt fragen, ob es nicht möglich sei, auch noch eine dritte der Constanten $s_{11}, \cdots, s_{33}$ durch geeignete Wahl der Coordinaten zum Verchwinden zu bringen. Da α_1, β_1, γ_1 bereits bestimmt sind, also die Lage der x-Achse festgelegt ist, so steht nur noch eine Drehung um die x-Achse frei, man hat demnach

$$x = \xi, \qquad y = \eta\cos\vartheta - \zeta\sin\vartheta, \qquad z = \eta\sin\vartheta + \zeta\cos\vartheta$$

zu setzen. Bildet man nach den vorher angegebenen Formeln die Grössen $\sigma_{11}, \cdots, \sigma_{33}$, so wird zunächst

$$\sigma_{11} = s_{11}, \qquad \sigma_{12} = 0, \qquad \sigma_{13} = 0,$$

und man erhält weiter

$$\sigma_{21} = s_{21}\cos\vartheta + s_{31}\sin\vartheta, \qquad \sigma_{22} = s_{22}\cos^2\vartheta + (s_{23} + s_{32})\cos\vartheta\sin\vartheta + s_{33}\sin^2\vartheta,$$
$$\sigma_{23} = (s_{33} - s_{22})\sin\vartheta\cos\vartheta + s_{21}\cos^2\vartheta - s_{32}\sin^2\vartheta;$$
$$\sigma_{31} = -s_{21}\sin\vartheta + s_{31}\cos\vartheta, \qquad \sigma_{32} = (s_{33} - s_{22})\sin\vartheta\cos\vartheta - s_{23}\sin^2\vartheta + s_{32}\cos^2\vartheta,$$
$$\sigma_{33} = s_{22}\sin^2\vartheta - (s_{23} + s_{32})\sin\vartheta\cos\vartheta + s_{33}\cos^2\vartheta.$$

Wie sich sogleich zeigen wird, hat es nur einen Nutzen σ_{23} oder σ_{32} zum Ver-

schwinden zu bringen, dafür aber erhält man Gleichungen zweiten Grades in $\operatorname{tg}\vartheta$, bei denen die Realität der Wurzeln ungewiss bleibt. Deshalb ist es besser, mit der Bestimmung von ϑ zu warten, bis weitere Folgerungen aus dem Verschwinden von s_{12} und s_{13} gezogen worden sind.

8.

Weil C_1 eine eigentliche Raumcurve ist, kann sich die zugehörige x-Coordinate nicht auf eine Constante reducieren, mithin ist x eine wirkliche Function von u, die man gleich dem neuen Parameter u_1 setzen darf. Folglich ist es gestattet, von vornherein $f(u) = 1$ anzunehmen. Bei dieser Annahme erhält man, wenn noch berücksichtigt wird, dass $s_{12} = 0$, $s_{13} = 0$ ist, an Stelle der Gleichungen (B) die einfacheren Gleichungen:

$$(B') \qquad \begin{cases} \dfrac{hg' - gh'}{1 + g^2 + h^2} = s_{11}, \\ \dfrac{h'}{1 + g^2 + h^2} = s_{21} + s_{22}g + s_{23}h, \\ \dfrac{-g'}{1 + g^2 + h^2} = s_{31} + s_{32}g + s_{33}h, \end{cases}$$

und die Relation (D) geht über in:

$$(D') \qquad 0 = s_{11} + (s_{21} + s_{22}g + s_{23}h)g + (s_{31} + s_{32}g + s_{33}h)h.$$

Diese Relation differentiiere man nach u und setze in dem Ergebnis für g' und h' die aus den beiden letzten Gleichungen (B') folgenden Werte ein. Dann kommt:

$$(E') \quad 0 = (s_{23}g + s_{33}h)(s_{21} + s_{22}g + s_{23}h) - (s_{22}g + s_{32}h)(s_{31} + s_{32}g + s_{33}h).$$

Die Gleichung (E') ist keine algebraische Folge von (D'), denn sie kann daraus nicht durch Multiplication mit einem constanten Factor hervorgehen. Sie kann aber auch keine wirkliche Relation zwischen g und h sein, denn sonst beständen zwischen diesen beiden Functionen zwei Relationen und sie würden sich auf Constanten reducieren, mithin müsste die Curve C_1 eben, ja sogar eine gerade Linie sein. Folglich ist (E') eine Identität in g und h, und wenn man die rechte Seite als Polynom in g und h darstellt, verschwinden alle Coefficienten. Zunächst ist also

$$s_{23}s_{21} - s_{22}s_{31} = 0, \qquad s_{33}s_{21} - s_{32}s_{31} = 0,$$

und da jetzt die Determinante

$$S = s_{11}(s_{22}s_{33} - s_{23}s_{32})$$

ist, so folgt, dass

$$s_{21} = 0, \qquad s_{31} = 0$$

sein muss. Weiter ergibt sich aus dem Verschwinden des Coefficienten von gh, dass entweder

$$s_{32} = s_{23}$$

oder

$$s_{32} = -s_{23}$$

ist. In dem ersten Fall verschwinden auch die Coefficienten von g^2 und h^2, und (E') ist eine Identität. In dem zweiten bedingt das Verschwinden dieser Coefficienten, dass entweder $s_{23} = s_{32} = 0$ ist, was auf den ersten Fall zurückleitet, oder dass $s_{22} = s_{33} = 0$ ist, das ist aber unmöglich, da alsdann S gleich Null sein würde.

Damit ist folgendes Ergebnis gewonnen: wenn die Gleichungen (B') mit einander verträglich sein sollen, so müssen die Relationen bestehen:

$$s_{21} = 0, \qquad s_{31} = 0, \qquad s_{32} = s_{23}.$$

Die Gleichungen (B') haben also notwendig die Form:

$$(B'') \qquad \begin{cases} \dfrac{hg' - gh'}{1 + g^2 + h^2} = s_{11}, \\ \dfrac{h'}{1 + g^2 + h^2} = s_{22}g + s_{23}h, \\ \dfrac{-g'}{1 + g^2 + h^2} = s_{23}g + s_{33}h. \end{cases}$$

9.

Jetzt ist es Zeit, die vorher betrachtete Drehung um die x-Achse auszuführen. Für die Coefficienten $\sigma_{11}, \cdots, \sigma_{33}$ ergeben sich dann die Werte:

$$\sigma_{11} = s_{11}, \qquad \sigma_{12} = 0, \qquad \sigma_{13} = 0;$$

$$\sigma_{21} = 0, \qquad \sigma_{22} = s_{22}\cos^2\vartheta + 2s_{23}\cos\vartheta\sin\vartheta + s_{33}\sin^2\vartheta,$$

$$\sigma_{23} = (s_{33} - s_{22})\sin\vartheta\cos\vartheta + s_{23}(\cos^2\vartheta - \sin^2\vartheta);$$

$$\sigma_{31} = 0, \qquad \sigma_{32} = (s_{33} - s_{22})\sin\vartheta\cos\vartheta + s_{23}(\cos^2\vartheta - \sin^2\vartheta),$$

$$\sigma_{33} = s_{22}\sin^2\vartheta + 2s_{23}\cos\vartheta\sin\vartheta + s_{33}\cos^2\vartheta.$$

Mithin ist $\sigma_{23} = \sigma_{32}$ und diese beiden Coefficienten verschwinden, wenn der Drehungswinkel ϑ aus der Gleichung

$$\operatorname{tg} 2\vartheta = \frac{2s_{23}}{s_{22} - s_{33}}$$

bestimmt wird, was bei reellen Werten von $s_{11}, \cdots, s_{33}$ stets auf reelle Art möglich ist. Folglich gibt es stets mindestens ein (bei reellen Flächen reelles)

Coordinatensystem, bei dem sich das System der Gleichungen (B) in der einfachen Form darstellt:

$$(B^*) \qquad \begin{cases} \dfrac{hg' - gh'}{1 + g^2 + h^2} = s_{11}, \\ \dfrac{h'}{1 + g^2 + h^2} = s_{22} g, \\ \dfrac{-g'}{1 + g^2 + h^2} = s_{33} h; \end{cases}$$

hierin bedeuten s_{11}, s_{22}, s_{33} von Null verschiedene Constanten, die keiner weiteren Bedingung unterworfen sind.

Aus den beiden letzten Gleichungen (B^*) folgt durch Division

$$-\frac{h'}{g'} = \frac{s_{22} g}{s_{33} h},$$

also ist

$$s_{22} g g' + s_{33} h h' = 0,$$

und man hat die Integralgleichung:

$$s_{22} g^2 + s_{33} h^2 = \text{const.} \tag{8}$$

Nimmt man jetzt die erste der Gleichungen (B^*) hinzu, so ergibt sich durch Combination mit den beiden letzten die schon früher aufgestellte Relation (D), die nunmehr die Gestalt annimmt:

$$(D^*) \qquad s_{11} + s_{22} g^2 + s_{33} h^2 = 0.$$

Folglich besagt die erste der Gleichungen (D^*), dass die beiden letzten mit der Massgabe integriert werden sollen, dass die Constante in der Integralgleichung (8) den Wert

$$-s_{11}$$

erhält.

Zur Abkürzung möge

$$-s_{11} : s_{33} = A, \qquad -s_{22} : s_{33} = B$$

gesetzt werden. Dann geht die Gleichung (8) über in die Gleichung:

$$h = \sqrt{A + B g^2};$$

das Vorzeichen der Wurzel darf beliebig gewählt werden. Aus der letzten der Gleichungen (B^*) ergibt sich dann zur Bestimmung von g die Gleichung

$$g' = -s_{33}[1 + A + (1 + B) g^2] \sqrt{A + B g^2}. \tag{9}$$

Ganz entsprechende Gleichungen erhält man aus den Gleichungen (C) für die Functionen k, l, m. Es wird

$$k = 1, \qquad m = \sqrt{A + B l^2},$$

und l ist zu bestimmen aus der Gleichung

$$l' = -s_{33}[1 + A + (1 + B)l^2]\sqrt{A + Bl^2}; \tag{10}$$

das Vorzeichen der Quadratwurzel darf wieder beliebig gewählt werden.

10.

Bei der Bestimmung von q und l aus den Gleichungen (9) und (10) bedürfen die Fälle, dass zwei der Grössen s_{11}, s_{22}, s_{33}, einander gleich werden, also die Fälle:

$$\text{I.}\quad A = -1, \qquad \text{II.}\quad B = -1, \qquad \text{III.}\quad A = B$$

einer besonderen Behandlung, die vorweg erledigt werden soll.

I. Ist $A = -1$, so setze man $B = c^2$. Dann wird

$$q = \frac{1}{c}\cdot\frac{1}{\sqrt{1-u_1^2}}, \qquad h = \frac{u_1}{\sqrt{1-u_1^2}},$$

wenn zur Abkürzung

$$-s_{33}\frac{1+c^2}{c}u = u_1$$

gesetzt wird. Hieraus folgt für

$$u_1 = \sin p,$$

dass die Coordinaten x_1, y, z der Curve C_1 bis auf einen und denselben constanten Factor gleich

$$c\cdot\sin p, \qquad \pm p, \qquad \pm c\cdot\cos p$$

sind, wobei entweder die oberen oder die unteren Vorzeichen gelten. Da es genügt, von den einander ähnlichen Flächen der betrachteten Art je einen Repräsentanten anzugeben, ist es erlaubt, x_1, y_1, z_1 gleich diesen Ausdrücken setzen; auch darf man unbeschadet der Allgemeinheit das obere Vorzeichen wählen, indem über die Richtungen der y- und z-Achse noch verfügt wird. In entsprechender Weise erhält man für die Coordinaten x_2, y_2, z_2 der Curve C_2 die Ausdrücke

$$c\cdot\sin q, \qquad \pm q, \qquad \pm c\cdot\cos q,$$

in denen wiederum entweder die oberen oder die unteren Vorzeichen gelten. Jetzt hat man aber beide Vorzeichen zu berücksichtigen und findet so für die gesuchten Flächen die Gleichungen

$$x = c(\sin p + \sin q), \qquad y = p + \epsilon q, \qquad z = c(\cos p + \epsilon\cos q), \tag{11}$$

in denen $\epsilon = \pm 1$ zu setzen ist.

Für $\epsilon = +1$ ergibt die Elimination von p und q die Gleichung

$$(L) \qquad x : z = \operatorname{tg}\tfrac{1}{2}y,$$

für $\epsilon = -1$ aber

$$z : x = -\operatorname{tg} \tfrac{1}{2} y,$$

sodass man in beiden Fällen die *gemeine geradlinige Schraubenfläche* erhält.

II. Ist $B = -1$, so setze man $A = c^2$. Dann wird ähnlich wie bei I.:

$$x = p + \epsilon q, \qquad y = c(\cos p + \epsilon \cos q), \qquad z = c(\sin p + \sin q),$$

sodass man wiederum zu der gemeinen geradlinigen Schraubenfläche geführt wird.

III. Ist endlich $A = B$, so gelangt man sofort zu dem Fall I. zurück.

Mithin ergeben sich immer, wenn zwei der drei Grössen s_{11}, s_{22}, s_{33} einander gleich sind, die aus der gemeinen geradlinigen Schraubenfläche (L) durch Aehnlichkeitstransformationen hervorgehenden Flächen, die ebenfalls gemeine geradlinige Schraubenflächen sind. Da in den Gleichungen (11) die willkürliche Constante c auftritt, die in der Gleichung (L) fehlt, so *lässt sich eine jede gemeine geradlinige Schraubenfläche auf unendlich viele Arten durch die Translation von Raumcurven, nämlich von cylindrischen Schraubenlinien, erzeugen.* Im übrigen sei für diese Flächen auf die Arbeiten von S. Lie und auf G. Scheffers, *Einführung in die Theorie der Flächen*, Leipzig, 1902, S. 190 bis 191 sowie Göttinger Nachrichten, November, 1905, verwiesen.

11.

Unter der Voraussetzung, dass die Grössen s_{11}, s_{22}, s_{33} von einander verschieden sind, folgt aus der Gleichung (9) durch Ausführung der Quadratur:

$$\frac{1}{\sqrt{(1+A)(A-B)}} \operatorname{arc\,tg}\left(\sqrt{\frac{A-B}{1+A}}\,\frac{g}{\sqrt{A+Bg^2}}\right) = -s_{33}u,$$

und die entsprechende Gleichung gilt für l und v. Setzt man also

$$u_1 = -s_{33}\sqrt{(1+A)(A-B)}\,u, \qquad v_1 = -s_{33}\sqrt{(1+A)(A-B)}\,v,$$

so wird:

$$g = \frac{\sqrt{A(1+A)}\sin u_1}{\sqrt{(A-B)\cos^2 u_1 - (1+A)B\sin^2 u_1}},$$

$$h = \frac{\sqrt{A(A-B)}\cos u_1}{\sqrt{(A-B)\cos^2 u_1 - (1+A)B\sin^2 u_1}};$$

$$l = \frac{\sqrt{A(1+A)}\sin v_1}{\sqrt{(A-B)\cos^2 v_1 - (1+A)B\sin^2 v_1}},$$

$$m = \frac{\sqrt{A(A-B)}\cos v_1}{\sqrt{(A-B)\cos^2 v_1 - (1+A)B\sin^2 v_1}};$$

das Vorzeichen der Quadratwurzeln, die beziehungsweise in g und h, l und m auftreten, ist willkürlich, nur muss bei g und h dasselbe Vorzeichen genommen werden, und ebenso bei l und m. Da man aber über die Richtungen der y- und der z-Achse noch verfügt, darf man unbeschadet der Allgemeinheit das Vorzeichen der in g und h auftretenden Wurzel fixieren, während dann die in l und m auftretende Wurzel zweiwertig bleibt.

Aus den so gefundenen Ausdrücken für f, g, h; k, l, m ergeben sich durch Ausführung der Quadraturen, wenn zur Vereinfachung statt u_1 und v_1 wieder u und v geschrieben wird, für x, y, z selbst die Gleichungen:

$$x = u + v,$$

$$y = -\sqrt{\frac{1+A}{1+B}}\left[\log\frac{\sqrt{A(1+B)}\cos u + \sqrt{(A-B)\cos^2 u - (1+A)B\sin^2 u}}{\sqrt{A(1+B)} + \sqrt{A-B}}\right.$$
$$\left. + \epsilon \log\frac{\sqrt{A(1+B)}\cos v + \sqrt{(A-B)\cos^2 v - (1+A)B\sin^2 u}}{\sqrt{A(1+B)} + \sqrt{A-B}}\right],$$

$$z = \sqrt{\frac{A-B}{1+B}}\left[\arcsin\left(\sqrt{\frac{A(1+B)}{A-B}}\sin u\right) + \epsilon \arcsin\left(\sqrt{\frac{A(1+B)}{A-B}}\sin v\right)\right].$$

Wenn jetzt noch

$$\sqrt{\frac{1+B}{A-B}} = \operatorname{tg}\gamma, \qquad \sqrt{\frac{A(1+B)}{A-B}} = \lambda$$

gesetzt und eine Aehnlichkeitstransformation im Verhältnis $1 : \sin\gamma$ vorgenommen wird, so ergeben sich schliesslich für die gesuchten Flächen die Gleichungen:

$$(S)\quad \begin{cases} x = \sin\gamma\cdot(u+v), \\ y = -\left[\log\dfrac{\lambda\cos u + \sqrt{1-\lambda^2\sin^2 u}}{\lambda+1} + \epsilon\log\dfrac{\lambda\cos u + \sqrt{1-\lambda^2\sin^2 v}}{\lambda+1}\right], \\ z = \cos\gamma\cdot[\arcsin(\lambda\sin u) + \epsilon\arcsin(\lambda\sin v)], \end{cases}$$

in denen γ und λ willkürliche Constanten bedeuten.

12.

Es werde zuerst der Fall $\epsilon = -1$ betrachtet, in dem die Gleichungen (S) übergehen in die Gleichungen:

$$(S_1)\quad \begin{cases} x = \sin\gamma\cdot(u+v), \\ y = -\log(\lambda\cos u + \sqrt{1-\lambda^2\sin^2 u}) + \log(\lambda\cos v + \sqrt{1-\lambda^2\sin^2 v}), \\ z = \cos\gamma\cdot[\arcsin(\lambda\sin u) - \arcsin(\lambda\sin v)]. \end{cases}$$

Bildet man die Functionaldeterminante von x, y, z nach u, v, λ, so ergibt eine einfache Rechnung, dass sie identisch verschwindet, und das bedeutet, dass in der Gleichung, die sich aus den Gleichungen (S_1) durch Elimination von u und

v ergibt, also in der Gleichung der betrachteten Fläche, bezogen auf die cartesischen Coordinaten x, y, z, die Constante λ fehlt. Nun erhält man aber für $\lambda = 1$:

$$(M) \qquad \begin{cases} x = \sin\gamma \cdot (u + v), \\ y = -\log\cos u + \log\cos v, \\ z = \cos\gamma \cdot (u - v), \end{cases}$$

also die früher gefundene Parameterdarstellung der Scherkschen Minimalflächen:

$$(M^*) \qquad e^y = \frac{\cos(\rho x - rz)}{\cos(\rho x + rz)} \qquad (\rho = 1 : 2\sin\gamma, \quad r = 1 : 2\cos\gamma).$$

Man wird hieraus schliessen, dass die Gleichungen (S_1) ebenfalls eine Darstellung der Scherkschen Minimalfläche (M^*) liefern oder vielmehr unendlich viele Darstellungen, indem ja die Constante λ für die erzeugenden Curven eine wesentliche Constante ist. Mithin *lässt sich eine jede Scherksche Minimalfläche auf unendlich viele Arten durch die Translation von eigentlichen Raumcurven erzeugen, deren Bogenelement nicht verschwindet.*

Um aus den Gleichungen (S_1) die von λ freie Gleichung zwischen x, y, z wirklich herzustellen, verfährt man am einfachsten so. Wird in diesen Gleichungen

$$v = 0$$

gesetzt, so ergibt sich eine Curve oder vielmehr eine Schar von Curven auf derjenigen Fläche, die zu dem betreffenden Werte des Constanten γ gehört, da ja in den Gleichungen

$$(T_1) \qquad \begin{cases} x = \sin\gamma \cdot u, \\ y = -\log(\lambda\cos u + \sqrt{1 - \lambda^2\sin^2 u}) + \log(\lambda + 1), \\ z = \cos\gamma \cdot \arcsin(\lambda\sin u) \end{cases}$$

noch die willkürliche Constante λ vorkommt. Der Ort der durch die Gleichungen (T_1) dargestellten Curven ist also die Fläche (S_1), und man braucht daher nur aus den Gleichungen (T_1) u und λ zu eliminieren, um die von λ freie Gleichung zwischen x, y, z zu erhalten, die sich dann in der Form (M^*) ergibt.

13.

Für $\epsilon = +1$ ergeben sich die Gleichungen:

$$(S_2) \qquad \begin{cases} x = \sin\gamma \cdot (u + v), \\ y = -\log\left(\dfrac{\lambda\cos u + \sqrt{1 - \lambda^2\sin^2 u}}{\lambda + 1}\right) - \log\left(\dfrac{\lambda\cos v + \sqrt{1 - \lambda^2\sin^2 v}}{\lambda + 1}\right), \\ z = \cos\gamma \cdot [\arcsin(\lambda\sin u) + \arcsin(\lambda\sin v)], \end{cases}$$

bei denen die Functionaldeterminante von x, y, z nach u, v, λ NICHT identisch verschwindet. Ich hatte daher zuerst geglaubt, dass die Gleichungen (S_2) Flächen liefern würden, die in den Gleichungen (S_1) nicht enthalten sind; wie jedoch G. SCHEFFERS (Göttinger Nachrichten, November, 1905) gezeigt hat, ergeben sich auch hier nur die Scherkschen Flächen, für die man so zwei wesentlich verschiedene Arten von Erzeugung durch Translation erhält.

Um zunächst aus (S_2) eine Schar von Flächen zu erhalten, die nur von dem einen Parameter γ abhängt, übe man auf jede Fläche (S_2) eine Translation parallel der y-Achse aus, bei der y um den Betrag

$$\log \frac{\lambda - 1}{\lambda + 1}$$

wächst. Auf diese Weise findet man die Flächenschar:

$$(S_2^*) \quad \begin{cases} x = \sin\gamma \cdot (u + v), \\ y = -\log\left(\dfrac{\lambda\cos u + \sqrt{1-\lambda^2\sin^2 u}}{\sqrt{\lambda^2-1}}\right) - \log\left(\dfrac{\lambda\cos v + \sqrt{1-\lambda^2\sin^2 v}}{\sqrt{\lambda^2-1}}\right), \\ z = \cos\gamma \cdot [\arcsin(\lambda\sin u) + \arcsin(\lambda\sin v)], \end{cases}$$

bei der jetzt die Functionaldeterminante nach u, v, λ identisch verschwindet, sodass man bei Elimination von u und v eine von λ freie Gleichung erhält. Um diese Gleichung herzustellen, braucht man nur zu bedenken, dass die Gleichungen (S_2^*) dieselbe Fläche ergeben, wie auch λ gewählt werde. Setzt man in ihnen

$$u = v,$$

so erhält man die auf der Fläche liegenden Curven:

$$(T_2) \quad \begin{cases} x = \sin\gamma \cdot 2u, \\ y = -2\log\left(\dfrac{\lambda\cos u + \sqrt{1-\lambda^2\sin^2 u}}{\sqrt{\lambda^2-1}}\right), \\ z = \cos\gamma \cdot 2\arcsin(\lambda\sin u), \end{cases}$$

deren Ort wiederum die Fläche ist. Die Gleichungen (T_2) sind also auch eine Darstellung der Fläche, nämlich in den Parametern u und λ, und jetzt ergibt sich aus (T_2) durch Elimination dieser Parameter sofort die Gleichung

$$e^y = \frac{\cos(\rho x - rz + \frac{1}{2}\pi)}{\cos(\rho x + rz - \frac{1}{2}\pi)},$$

in der ρ und r dieselben Werte, wie bei (M^*) haben.

Das so gewonnene Ergebnis, dass *sich die Scherkschen Flächen auf zwei*

wesentlich verschiedene Arten durch Translation erzeugen lassen, steht in Einklang mit einem Satze von S. LIE, nach dem man die erzeugenden Curven auf folgende Art erhält. Aus einer der beiden Scharen von Asymptotenlinien (Haupttangentencurven) der Fläche greife man irgend eine Curve heraus, nehme einen ihrer Punkte und ziehe von diesem aus alle Sehnen: die Mitten dieser Sehnen bilden dann eine erzeugende Curve. Je nachdem die herausgegriffene Asymptotenlinie der einen oder der anderen Schar angehört, erhält man die eine oder die andere der beiden Erzeugungsarten, und zwar wird durch die Gleichungen (T_2) die eine der beiden Scharen, durch die Gleichungen (S_2^*) die zugehörige Erzeugung dargestellt.

HANNOVER, TECHNISCHE HOCHSCHULE.

A FIFTH NECESSARY CONDITION FOR A STRONG EXTREMUM OF THE INTEGRAL

$$\int_{x_0}^{x_1} F(x, y, y')\,dx^*$$

BY

OSKAR BOLZA

In a previous article† I called attention to the fact that *the conditions of Euler, Legendre, Jacobi and Weierstrass are not sufficient* for a strong extremum of the definite integral

$$J = \int_{x_0}^{x_1} F(x, y, y')\,dx.$$

Hence the question of further necessary conditions arises, and the object of the present paper is to derive *a fifth necessary condition.*

§ 1. *Preliminary form of condition* (V).

The terminology, and the assumptions concerning the function $F(x, y, y')$ and the admissible curves being the same as in § 3, c) of my *Lectures on the Calculus of Variations*, let

$$\mathfrak{E}_0: \qquad y = f_0(x), \qquad x_0 \leqq x \leqq x_1,$$

be an extremal of class C' which passes through the two given points $P_0(x_0, y_0)$ and $P_1(x_1, y_1)$ and which lies in the interior of the region **R** to which the admissible curves are confined. We suppose that for the curve $\mathfrak{E}_0$ the conditions of LEGENDRE and JACOBI are fulfilled in the somewhat stronger form

(II′) $$F_{y'y'}\big(x, f_0(x), f_0'(x)\big) > 0 \text{ in } (x_0 x_1),$$

(III′) $$x_1 < x_0',$$

x_0' being the conjugate value to x_0.

* Presented to the Society February 24, 1906. Received for publication January 2, 1906.

† *Some Instructive Examples in the Calculus of Variations*, Bulletin of the American Mathematical Society, vol. 9 (1902), p. 9. Compare also my *Lectures on the Calculus of Variations* (Chicago, 1904), p. 99.

We proceed at first exactly as in one of the proofs* for the necessity of WEIERSTRASS' condition:

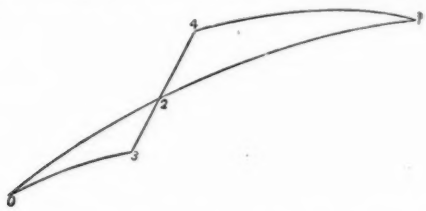
From (II′) and (III′) it follows that the set of extremals through the point P_0:

$$y = \overset{0}{\phi}(x, \gamma) \tag{1}$$

furnishes an "improper" field† $\mathbf{S}$ about the arc $\mathfrak{E}_0$. Let now $P_2(x_2, y_2)$ be any point of $\mathfrak{E}_0$ to the right of P_0, i. e., $x_0 < x_2 \leqq x_1$, and take ρ so small that the neighborhood (ρ) of P_2 lies at the same time in the interior of the field $\mathbf{S}$ and of the region $\mathbf{R}$. Then if we choose $0 < h < \rho$, $|k| < \rho$ and denote by P_3 the point whose coördinates are

$$x_3 = x_2 - h, \qquad y_3 = y_2 - k,$$

there passes one and but one extremal P_0P_3 of the field through the point P_3. We draw the straight line P_3P_2 and vary the extremal $\mathfrak{E}_0$ by replacing the arc



P_0P_2 by the broken curve $P_0P_3P_2$. To this variation we may apply WEIERSTRASS' theorem and obtain, since the $\mathbf{E}$-function vanishes along the extremal P_0P_3, for the total variation of J the expression

$$\Delta J = \int_0^1 h\mathbf{E}\left(x, \bar{y}; \overset{0}{p}(x, \bar{y}), \frac{k}{h}\right) dt, \tag{2}$$

where

$$x = x_2 - ht, \qquad \bar{y} = y_2 - kt,$$

and $\overset{0}{p}(x, \bar{y})$ denotes the slope at the point $(x, \bar{y})$ of the unique extremal of the field passing through the point $(x, \bar{y})$.

If we keep the line P_3P_2 (i. e., the ratio k/h) fixed and let the point P_3 approach the point P_2 along this line, we obtain WEIERSTRASS' condition. But if, on the contrary, we *revolve the line P_3P_2 about P_2 and let it approach the position parallel to the y-axis*, while the point P_3 moves on a line parallel to the x-axis (i. e., if we keep k fixed and let h approach zero), we obtain a new

* Compare E. R. HEDRICK, *On the sufficient conditions in the Calculus of Variations*, Bulletin of the American Mathematical Society, vol. 9 (1902), p. 14.

† Compare my *Lectures*, pp. 61 and 83, footnote 2; further KNESER, *Lehrbuch der Variationsrechnung*, § 27 and GOURSAT, *Cours d' Analyse*, vol. 2, p. 615, when the function F is analytic, and LUNN, in a paper which will be published in these Transactions, when F is not analytic.

necessary condition. For if the extremal $\mathfrak{E}_0$ furnishes a strong minimum for the integral J, the lower bound* of ΔJ for $h = +0$ must be positive or zero for all sufficiently small values of $|k|$.

The same reasoning can be applied to the set of extremals through the point P_1:

$$y = \overset{1}{\phi}(x, \gamma) \tag{3}$$

and the variation $P_2 P_4 P_1$ (see figure). If we introduce the symbols

$$\epsilon_0 = -1, \qquad \epsilon_1 = +1,$$

the results of the two processes may be united in the one formula

$$\lim_{h=+0} \int_0^1 h\mathbf{E}\left(x, \bar{y}; \overset{\lambda}{p}(x, \bar{y}), \frac{k}{h}\right) dt \geqq 0 \qquad (\lambda = 0, 1), \tag{4}$$

where

$$x = x_2 + \epsilon_\lambda ht, \qquad \bar{y} = y_2 + \epsilon_\lambda kt,$$

and $\overset{\lambda}{p}(x, \bar{y})$ denotes the slope at the point $(x, \bar{y})$ of the unique extremal passing through the points P_λ and $(x, \bar{y})$.

If we substitute in (4) for the **E**-function its explicit expression

$$\mathbf{E}(x, y; p, \tilde{p}) = F(x, y, \tilde{p}) - F(x, y, p) - (\tilde{p} - p)F_{y'}(x, y, p),$$

the expression for ΔJ breaks up into four definite integrals two of which approach zero for $h = +0$, and we obtain the following *preliminary form of condition* (V):

$$\lim_{h=+0} \int_0^1 hF\left(x_2 + \epsilon_\lambda ht, y_2 + \epsilon_\lambda kt, \frac{k}{h}\right) dt$$
$$- k\int_0^1 F_{y'}\left(x_2, y_2 + \epsilon_\lambda kt, \overset{\lambda}{p}(x_2, y_2 + \epsilon_\lambda kt)\right) dt \geqq 0, \tag{5}$$

($\lambda = 0, 1$ when $x_0 < x_2 < x_1$; $\lambda = 0$ when $x_2 = x_1$; $\lambda = 1$ when $x_2 = x_0$).

These two inequalities must be satisfied for every value of k — positive or negative — of sufficiently small absolute value.

§ 2. *Final form of condition* (V).

From our assumptions concerning the function F and the properties† of the slope $\overset{\lambda}{p}(x, y)$ it follows by applying TAYLOR's formula that

$$\int_0^1 F_{y'}\left(x_2, y_2 + \epsilon_\lambda kt, \overset{\lambda}{p}(x_2, y_2 + \epsilon_\lambda kt)\right) dt = \overset{\lambda}{X}(x_2, y_2) + \frac{\epsilon_\lambda k}{2} \overset{\lambda}{X}_y(x_2, y_2) + k(k),$$

* "Untere Unbestimmtheitsgrenze" = "Unterer Limes," compare *Encyclopaedie*, II, A 1 (PRINGSHEIM), p. 14.

† Compare my *Lectures*, pp. 81, 82.

where (k) is an infinitesimal for $\mathbf{L}k = 0$, and

$$\overset{\lambda}{X}(x, y) = F_{y'}(x, y, \overset{\lambda}{p}(x, y)). \tag{6}$$

Since the point P_2 lies on the extremal $\mathfrak{E}_0$,

$$\overset{\lambda}{p}(x_2, y_2) = f'_0(x_2) \equiv y'_2, \tag{7}$$

and therefore

$$\overset{\lambda}{X}(x_2, y_2) = F_{y'}(x_2, y_2, y'_2). \tag{8}$$

Again,

$$\overset{\lambda}{X}_y(x, y) = F_{y'y}\big(x, y, \overset{\lambda}{p}(x, y)\big) + F_{y'y'}\big(x, y, \overset{\lambda}{p}(x, y)\big)\overset{\lambda}{p}_y(x, y)$$

and *

$$\overset{\lambda}{p}_y(x, y) = \frac{\overset{\lambda}{\phi}_{x\gamma}(x, \gamma)}{\overset{\lambda}{\phi}_{\gamma}(x, \gamma)},$$

γ being replaced by the inverse function: $\gamma = \overset{\lambda}{\psi}(x, y)$ obtained by solving the equation: $y = \overset{\lambda}{\phi}(x, y)$ with respect to γ.

If we substitute in these formulas x_2, y_2 for x, y, we must give γ the particular value γ_0 which corresponds to the extremal $\mathfrak{E}_0$ since P_2 lies on $\mathfrak{E}_0$. Hence if we write for brevity

$$\begin{aligned} F_{y'y}\big(x, f_0(x), f'_0(x)\big) &= Q(x), \\ F_{y'y'}\big(x, f_0(x), f'_0(x)\big) &= R(x), \end{aligned} \tag{9}$$

we obtain

$$\overset{\lambda}{X}_y(x_2, y_2) = Q(x_2) + R(x_2)\frac{\overset{\lambda}{\phi}_{x\gamma}(x_2, \gamma_0)}{\overset{\lambda}{\phi}_{\gamma}(x_2, \gamma_0)}.$$

This expression may still be thrown into a different form by introducing the general solution

$$y = f(x, \alpha, \beta)$$

of EULER's differential equation. For if we denote by α_0, β_0, the special values of the constants of integration α, β, which furnish the extremal $\mathfrak{E}_0$ and put

$$\Delta(x, x_\lambda) = f_\alpha(x, \alpha_0, \beta_0)f_\beta(x_\lambda, \alpha_0, \beta_0) - f_\beta(x, \alpha_0, \beta_0)f_\alpha(x_\lambda, \alpha_0, \beta_0), \tag{10}$$

then †

$$\overset{\lambda}{\phi}_{\gamma}(x, \gamma_0) = C_\lambda \Delta(x, x_\lambda),$$

where C_λ is a constant different from zero. Hence

$$\overset{\lambda}{X}_y(x_2, y_2) = Q(x_2) + R(x_2)\frac{\Delta_x(x, x_\lambda)}{\Delta(x, x_\lambda)}. \tag{11}$$

* Compare my *Lectures*, pp. 81, 82.

† Compare, for instance, my *Lectures*, p. 62.

Substituting the values of $\hat{X}(x_2, y_2)$ and $\hat{X}_{y'}(x_2, y_2)$ in (5), our condition may be written in the following form:

$$\text{(V)} \qquad \begin{aligned} &\mathop{\mathbf{L}}_{k=+0} \int_0^1 hF\left(x_2 + \epsilon_\lambda ht,\ y_2 + \epsilon_\lambda kt,\ \frac{k}{h}\right) dt - kF_{y'}(x_2, y_2, y_2') \\ &\qquad - \frac{\epsilon_\lambda k^2}{2}\left(Q(x_2) + R(x_2)\frac{\Delta_x(x, x_\lambda)}{\Delta(x, x_\lambda)}\right) + k^2(k) \geqq 0, \end{aligned}$$

in which it is immediately applicable to examples.

We now divide (V) by k and then let k approach zero. If we put

$$(12) \qquad U_\lambda(k, x_2) = \frac{1}{k} \mathop{\mathbf{L}}_{k=+0} \int_0^1 hF\left(x_2 + \epsilon_\lambda ht,\ y_2 + \epsilon_\lambda kt,\ \frac{k}{h}\right) dt$$

and

$$(13) \qquad \mathop{\mathbf{L}}_{k=+0} U_\lambda(k, x_2) = \overset{+}{G}_\lambda(x_2), \qquad \mathop{\mathbf{L}}_{k=-0} U_\lambda(k, x_2) = \overset{-}{G}_\lambda(x_2),$$

where $+\infty$ and $-\infty$ must be included among the possible values of $U_\lambda(k, x_2)$, and $\overset{\pm}{G}_\lambda(x_2)$, we reach the following

THEOREM: *In order that the extremal $\mathfrak{E}_0$, for which conditions* (II′) *and* (III′) *are supposed to be satisfied, may furnish a strong minimum for the integral*

$$J = \int_{x_0}^{x_1} F(x, y, y')\, dx,$$

it is necessary that

$$(\mathrm{V}_a) \qquad \begin{aligned} \overset{+}{G}_\lambda(x_2) - F_{y'}(x_2, y_2, y_2') &\geqq 0 \\ \overset{-}{G}_\lambda(x_2) - F_{y'}(x_2, y_2, y_2') &\leqq 0 \end{aligned}$$

for $\lambda = 0$ and for $\lambda = 1$, when $x_0 < x_2 < x_1$; for $\lambda = 0$, when $x_2 = x_1$; for $\lambda = 1$, when $x_2 = x_0$.

If the four (or two, when $x_2 = x_0$ or x_1) inequalities (V_a) are satisfied *with the inequality sign*, then also (V) holds for all sufficiently small values of $|k|$.

But it may happen that one or several of these inequalities hold *with the equality sign.* In this case we cannot go back to (V), unless a further condition be added. We obtain it by dividing (V) by k^2 before passing to the limit. If we put

$$(14) \qquad \mathop{\mathbf{L}}_{k=\pm 0} \frac{U_\lambda(k, x_2) - F_{y'}(x_2, y_2, y_2')}{k} = \overset{\pm}{H}_\lambda(x_2),$$

including again the values $+\infty$ and $-\infty$ among the possible values of $\overset{\pm}{H}_\lambda(x_2)$, we obtain the following

COROLLARY: *If one of the conditions* (V_a) *is satisfied in the form of an equality*

$$(15) \qquad \overset{\pm}{G}_\lambda(x_2) - F_{y'}(x_2, y_2, y_2') = 0,$$

then, with the same meaning of $\pm$ *and* λ, *the following additional condition must hold* *

$$(V_b) \qquad \overset{*}{H}_\lambda(x_2) - \frac{\epsilon_\lambda}{2} Q(x_2) - \frac{\epsilon_\lambda}{2} R(x_2) \frac{\Delta_1(x_2, x_\lambda)}{\Delta(x_2, x_\lambda)} \geqq 0,$$

where the functions $Q(x)$, $R(x)$, $\Delta(x, x_\lambda)$ *are defined by* (9) *and* (10).

If condition (V_b) is satisfied with the equality sign, another condition must be added, derived from the terms of the third order in the expansion of (V), etc., etc.

§ 3. *The special case when* $F(x, y, p)$ *admits expansions into power series in the vicinity of* $p = \pm \infty$.

The lower bounds U, G, H can be easily computed when $F(x, y, p)$ admits expansions into power series of the form

$$(16) \qquad F(x, y, p) = p^{n_1} \sum_{i,j,l} A^{(1)}_{i,j,l} (x - x_2)^i (y - y_2)^j \left(\frac{1}{p}\right)^l$$

convergent for $|x - x_2| < d_1$, $|y - y_2| < d_1$, $p > R_1$, and

$$(17) \qquad F(x, y, p) = |p^{n_2}| \sum_{i,j,l} A^{(2)}_{i,j,l} (x - x_2)^i (y - y_2)^j \left(\frac{1}{p}\right)^l$$

convergent for $|x - x_2| < d_2$, $|y - y_2| < d_2$, $p < -R_2$, d_1, d_2, R_1, R_2 being positive and the indices i, j, l running from 0 to $+\infty$; n_1 and n_2 are real, but need not be integers.

Under these assumptions we find easily

$$\int_0^1 hF\left(x_2 + \epsilon ht, y_2 + \epsilon kt, \frac{k}{h}\right) dt = \frac{|k|^n}{h^{n-1}} \sum_{i,j,l} \frac{\epsilon^{i+j}}{i+j+1} A_{ijl} k^{i+j} \left(\frac{h}{k}\right)^{l+i},$$

where n and A_{ijl} have to be replaced by n_1 and $A^{(1)}_{ijl}$ or by n_2 and $A^{(2)}_{ijl}$ according as k is positive or negative. Hence if we put

$$(18) \qquad B_{\mu\nu} = \sum_{i=0}^{\infty} A_{i,\mu-i,\nu-i},$$

with the understanding that every A_{ijl} with a negative index is zero, and suppose that

$$B_{\mu\nu} = 0 \qquad \text{for} \qquad \begin{cases} \mu = 0, 1, 2, \cdots, \\ \nu = 0, 1, 2, \cdots, s-1, \end{cases}$$

* If we consider the set of extremals through an arbitrary point $P(\xi, \eta)$ of $\mathfrak{E}_0$ different from P_2 we obtain a condition analogous to (V_b), in which x_λ is replaced by ξ and ϵ_λ by the sign of $(\xi - x_2)$. If we reduce the left-hand side to a fraction with the denominator $\Delta(x_2, \xi)$, numerator and denominator considered as functions of ξ are solutions of JACOBI's differential equation. Applying STURM's oscillation-theorem we obtain the result that the inequality in question is satisfied for every $\xi \neq x_2$ between x_0 and x_1, whenever it is satisfied for $\xi = x_0$ and $\xi = x_1$, so that no new condition is reached by this apparent generalization.

but that not all the coefficients $B_{\mu s}$ are equal to zero, we get for the above integral the value

$$|k|^{n-s}\delta^{r}\left\{\sum_{\mu=0}^{r}\frac{\epsilon^{\mu}}{\mu+1}B_{\mu s}k^{\mu}+(h)\right\},$$

where (h) is an infinitesimal and $\delta = k/|k|$.

We have now to distinguish three cases:

Case I. $n-s-1>0$.

Suppose $B_{\mu s}=0$ for $\mu=0,1,\cdots,r-1$, but $B_{rs}\neq 0$, then *condition* (V) *reduces to the inequality*

$$\epsilon_{\lambda}^{r}\delta^{r+s}B_{rs}>0. \tag{19}$$

If $x_0<x_2<x_1$, (19) must hold for $\epsilon_\lambda=+1$ and for $\epsilon_\lambda=-1$, hence r *must be even*. If $x_2=x_0$, (19) must hold for $\epsilon_\lambda=+1$, if $x_2=x_1$, for $\epsilon_\lambda=-1$.

Case II. $n-s-1=0$.

In this case the expansion of the left-hand side of (V) according to powers of k begins with the terms

$$k\left[\delta^{s+1}B_{0s}-F_{y'}(x_2,y_2,y_2')\right]+\frac{\epsilon_\lambda k^2}{2}\left[\delta^{s+1}B_{1s}-\overset{\lambda}{X}_{y}(x_2,y_2)\right]+\cdots.$$

Hence it follows that *condition* (V) *reduces to the inequality*

$$\delta\left[\delta^{s+1}B_{0s}-F_{y'}(x_2,y_2,y_2')\right]\geqq 0, \tag{20}$$

and *if*

$$\delta^{s+1}B_{0s}-F_{y'}(x_2,y_2,y_2')=0, \tag{21}$$

the condition must be added

$$\epsilon_\lambda\left(\delta^{s+1}B_{1s}-Q(x_2)-R(x_2)\frac{\Delta_x(x_2,x_\lambda)}{\Delta(x_2,x_\lambda)}\right)\geqq 0. \tag{22}$$

Case III. $n-s-1<0$.

In this case $U_\lambda(k,x_2)=0$ for $\lambda=0$, 1 and *condition* (V) *reduces to*

$$-\delta F_{y'}(x_2,y_2,y_2')\geqq 0 \tag{23}$$

and *if*

$$F_{y'}(x_2,y_2,y_2')=0 \tag{24}$$

the condition must be added

$$-\epsilon_\lambda\left(Q(x_2)+R(x_2)\frac{\Delta_x(x_2,x_\lambda)}{\Delta(x_2,x_\lambda)}\right)\geqq 0. \tag{25}$$

If $x_0<x_2<x_1$, conditions (22) and (24) must be satisfied for $\lambda=0$ *and* for $\lambda=1$; if $x_2=x_0$, for $\lambda=1$; if $x_2=x_1$, for $\lambda=0$.

In applying these conditions we must, in general, treat separately the two cases where k is positive ($n=n_1$, $A_{ijl}=A^{(1)}_{ijl}$, $\delta=+1$) and where k is negative ($n=n_2$, $A_{ijl}=A^{(2)}_{ijl}$, $\delta=-1$).

But when n_1 is an integer and *the same expansion holds for positive and negative values of* p — which happens, for instance, whenever $F(x, y, p)$ is a *rational function of* p — we have

$$n_1 = n_2, \text{ say } = n,$$

$A^{(2)}_{ijl} = (-1)^{n_1} A^{(1)}_{ijl}$, and therefore

$$r_1 = r_2, \text{ say } = r; \qquad s_1 = s_2, \text{ say } = s.$$

Hence it follows that the conditions for positive and negative values of k combine in the following manner:

Case I. $n - s - 1 > 0$,

(26) $\qquad n + r + s$ must be *even*, $\qquad \epsilon_\lambda^r B^{(1)}_{r_\alpha} > 0$.

Case II. $n - s - 1 = 0$,

(27) $$B^{(1)}_{0_\alpha} - F_{y'}(x_2, y_2, y_2') = 0,$$

(28) $$\epsilon_\lambda \left(B^{(1)}_{1_\alpha} - Q(x_2) - R(x_2) \frac{\Delta_\alpha(x_2, x_\lambda)}{\Delta(x_2, x_\lambda)} \right) \geqq 0.$$

Case III. $n + s - 1 < 0$.

Same conditions as in case II with $B^{(1)}_{0_\alpha} = 0$, $B^{(1)}_{1_\alpha} = 0$.

§ 4. *Examples.*

In this last section we propose to give examples for the different cases discussed in § 3. They will be so selected that not only conditions (II′) and (III′) are satisfied, but also WEIERSTRASS' condition in the somewhat stronger form

(IV′) $$\mathbf{E}\left(x, f_0(x); f_0'(x), \tilde{p}\right) > 0$$

for $x_0 \leqq x \leqq x_1$, and for every finite $\tilde{p} \neq f_0'(x)$, in order to show at the same time that our condition is *independent of* WEIERSTRASS' *condition*.

Case I. To this case belongs the example which I have given in the article referred to in the introduction:

$$F = ay'^2 - 4byy'^3 + 2bxy'^4,$$

$$a > 0, \qquad b > 0, \qquad (x_0, y_0) = 0, \qquad (x_1, y_1) = (1, 0).$$

The extremals are straight lines; in particular $\mathfrak{E}_0$ is the segment of the x-axis between $x = 0$ and $x = 1$. Conditions (II′), (III′) and (IV′) are satisfied Further we find easily for the integral

$$S = \int_0^1 hF\left(x_2 + \epsilon_\lambda ht, y_2 + \epsilon_\lambda kt, \frac{k}{h}\right) dt$$

the value

$$S = \frac{k^2}{h^3}(2bk^2 x_2 - bk^2 \epsilon_\lambda h + ah^2),$$

since $y_2 = 0$.

If $0 < x_2$ we obtain $U_\lambda(k, x_2) = +\infty$, for $\lambda = 0, 1$, and condition (V) is satisfied.

If $x_2 = 0$, in which case (V) must be satisfied for $\epsilon_\lambda = +1$, the term $-bk^4/h^2$ determines the sign and therefore $U_1(k, 0) = -\infty$, and (V) is not satisfied. Hence there exists *no strong minimum* if, as we suppose, the interval of integration extends to the point $x = 0$.

Also the following example, due to CARATHEODORY,* belongs to this case, viz.,

$$F = y'^2 - y^2 y'^4.$$

Here the extremals are, in general, not straight lines; but the particular line

$$\mathfrak{E}_0: \qquad y = 0 \equiv f_0(x)$$

is an extremal. Since

$$F_{y'y'}\big(x, f_0(x), f_0'(x)\big) = 2$$

condition (II′) is fulfilled, and we can always take x_0 and x_1 so close together that also (III′) is fulfilled. Again

$$\mathbf{E}\big(x, f_0(x), f_0'(x), \tilde{p}\big) = \tilde{p}^2;$$

hence also (IV′) is fulfilled.

Nevertheless there exists *no strong minimum*. For

$$S = \frac{k^2}{h^3}\left(-\frac{k^4}{3} + h^2 k^2\right),$$

and therefore $U_\lambda(k, x_2) = -\infty$ for $\lambda = 0, 1$ and for every x_2.

Case II. The examples for the remaining two cases will be of the form

$$F = \frac{L + My' + Ny'^2}{(1 + y'^2)^n},$$

where L, M, N are functions of x and y. We choose them so that all the non-singular extremals are straight lines, i. e., so that

$$F_y - F_{y'x} - y' F_{y'y} \equiv 0.$$

This leads to the following partial differential equations

$$\frac{\partial L}{\partial y} - \frac{\partial M}{\partial x} = 0, \qquad n\frac{\partial L}{\partial x} - \frac{\partial N}{\partial x} = 0,$$

$$(29) \qquad (2n+1)\frac{\partial L}{\partial y} + (2n-1)\frac{\partial M}{\partial x} - \frac{\partial N}{\partial y} = 0,$$

$$n\frac{\partial M}{\partial y} + (n-1)\frac{\partial N}{\partial x} = 0, \qquad (2n-1)\frac{\partial N}{\partial y} = 0.$$

*Archiv der Mathematik und Physik, ser. 3, vol. 10 (1906).

For $n \neq 0, 1, \frac{1}{2}$, the most general solution of these differential equations is

$$L = ax + b, \qquad M = -(n-1)ay + c, \qquad N = nax + d, \tag{30}$$

$a, b, \cdots$ being arbitrary constants.

For $n = \frac{1}{2}$, the most general solution is

$$\begin{aligned} L &= 2ax^2 + ay^2 + 2dx + by + c, \\ M &= 2axy + bx + dy + e, \\ N &= ax^2 + 2ay^2 + dx + 2by + f. \end{aligned} \tag{31}$$

The extremals being straight lines, condition (III′) is always satisfied.

We take the two given points on the x-axis, so that

$$\mathfrak{E}_0: \qquad y = 0 \equiv f_0(x).$$

Then

$$\begin{aligned} F_{y'}\big(x, f_0(x), f_0'(x)\big) &= M(x, 0), \qquad Q(x) = M_y(x, 0), \\ R(x) &= 2\big(N(x, 0) - nL(x, 0)\big). \end{aligned} \tag{32}$$

In order to obtain an example for case II, we choose $n = \frac{1}{2}$, and give the arbitrary constants in (31) the values $a = 0, b = 1, c = 1, d = 0, e = 0, f = 1$, so that

$$F = \frac{(y+1) + xy' + (2y+1)y'^2}{\sqrt{1 + y'^2}},$$

the square root being positive. Then

$$R(x) = +1,$$

so that (II′) is satisfied. Again, if we put $\tilde{p} = \tan\tilde{\theta}$, where $-\pi/2 < \tilde{\theta} < +\pi/2$, we get

$$\mathbf{E}\big(x, f_0(x); f_0'(x), \tilde{p}\big) = \frac{1 - \cos\tilde{\theta}}{\cos\tilde{\theta}}(1 - x\sin\tilde{\theta}).$$

Hence condition (IV′) will be satisfied if

$$-1 \leqq x_0 < x_1 \leqq 1. \tag{33}$$

For the integral S we find easily

$$S = \frac{k^2(\epsilon_\lambda k + 1) + hk\left(x_2 + \epsilon_\lambda \frac{h}{2}\right) + h^2\left(\epsilon_\lambda \frac{k}{2} + 1\right)}{\sqrt{h^2 + k^2}};$$

Further

$$F_{y'}[x, f_0(x), f_0'(x)] = x, \qquad Q(x) = 0, \qquad \Delta(x, x_\lambda) = x - x_\lambda.$$

Hence the inequality (V) takes the form

$$(\delta - x_2)k + \epsilon_\lambda k^2\Big(\delta - \frac{1}{2(x_2 - x_\lambda)}\Big) + k^2(k) \geqq 0,$$

where again $\delta = k/|k|$.

The discussion of this inequality leads to the following result:

When $-1 < x_0 < x_1 < +1$ *condition* (V) *is always fulfilled. But when* $x_1 = +1$, *or* $x_0 = -1$, *condition* (V) *introduces a new restriction of the interval beyond the restriction* (33) *already introduced by* WEIERSTRASS' *condition* (IV'), *viz., when* $x_1 = +1$, *we must have* $x_0 \geqq \frac{1}{2}$, *and when* $x_0 = -1$, *we must have* $x_1 \leqq \frac{1}{2}$.

Case III. In order to obtain an example for this case, we take $n = 2$, and give the constants in (30) the values $a = -1$, $b = 0$, $c = 0$, $d = 1$, so that

$$F = \frac{-x + yy' + (1 - 2x)y'^2}{(1 + y'^2)^2}.$$

Here we find

$$R(x) = 2,$$

so that condition (II') is satisfied. Further

$$\mathbf{E}\left(x, f_0(x); f_0'(x), \tilde{p}\right) = \frac{\tilde{p}^2(1 + x\tilde{p}^2)}{(1 + \tilde{p}^2)^2};$$

hence also (IV') *is satisfied provided that*

$$0 \leqq x_0.$$

We see at once that

$$\lim_{k=+0} S = 0.$$

And since

$$F_{y'}[x, f_0(x), f_0'(x)] \equiv 0, \qquad Q(x) = 1, \qquad \Delta(x, x_\lambda) = x - x_\lambda,$$

condition (V) reduces to

$$-\epsilon_\lambda\Big(1 + \frac{2}{x_2 - x_\lambda}\Big) \geqq 0.$$

The discussion of this inequality easily leads to the following result: *In the present example condition* (V) *is equivalent to a restriction of the length of the interval* $(x_0 x_1)$, *viz:*

$$x_1 - x_0 \leqq 2.$$

FREIBURG i. B.,
December 14, 1905.

A PROBLEM OF THE CALCULUS OF VARIATIONS IN WHICH THE INTEGRAND IS DISCONTINUOUS*

BY

GILBERT AMES BLISS AND MAX MASON

In the calculus of variations the assumption is made, in dealing with integrals of the form

$$J = \int F(x, y, x', y')dt,$$

that F is a continuous function of its four arguments. In the following pages the properties of a minimizing curve,

$$x = \phi(t), \qquad y = \psi(t), \tag{1}$$

for the integral J will be investigated when the function F, in general continuous, is allowed to have a finite discontinuity when the point (x, y) passes through a given curve.

This case would arise in the treatment by calculus of variations of the physical phenomena which take place in a medium whose properties change abruptly. For example, a ray of light follows such a path from one point, 1, to another, 2, that the time of passage is the shortest possible.† If we take the points 1, 2, in the xy-plane and assume the medium to be such that the velocity of light is given by a function $V(x, y)$ independent of z, then the path of the ray is a plane curve of the form (1) which minimizes the integral

$$\int_1^2 \frac{\sqrt{x'^2 + y'^2}}{V(x, y)}\, dt.$$

If abrupt changes in V occur, i. e., if refracting surfaces are present, the integrand has discontinuities of the kind above mentioned.

The problem to be treated in the following pages may be stated as follows: *Among all curves which connect two given points* 1 *and* 3 *lying on opposite sides of a fixed curve* L, *and which cross* L *but once, that one* C *is required which minimizes the sum of the two integrals*

$$J = \int F(x, y, x', y')dt,$$

$$j = \int f(x, y, x', y')dt,$$

* Presented to the Society October 28, 1905. Received for publication December 18, 1905.

† In some cases the time is a maximum, but in a sense different from the one ordinarily understood in the calculus of variations. See, e. g., DRUDE, *Lehrbuch der Optik*, p. 11.

where the first integral is to be taken from the fixed point 1 *to the curve* L, *the second from* L *to the fixed point* 3.

The intersection of the curve L with the solution C will be denoted by 2, the intersection of L with any comparison curve by $2'$.

Three of the necessary conditions which the curve C_{12} must satisfy in order to minimize the integral are readily derivable from the known conditions in the ordinary problem of the calculus of variations. In § 2 a fourth condition corresponding to the Jacobi condition will be derived, and it will be shown in § 3 that the four conditions with the usual slight alterations are also sufficient.

§ 1. *The known necessary conditions.*

All the curves considered are supposed to lie in a closed region R of the xy-plane. This region is divided by the curve L,

$$L: \qquad x = a(u), \qquad y = b(u)$$

into two closed regions which contain the points 1 and 3 respectively. The two functions F and f in the integrals J and j will be assumed to have similar properties over the entire region R.* These properties need be stated for only one of them. In the region R the function $F(x, y, x', y')$ is of class C'''† for all values of x' and y' not both zero. F has also the usual homogeneity property ‡

$$F(x, y, \kappa x', \kappa y') = \kappa F(x, y, x', y') \qquad (\kappa > 0), \tag{2}$$

or

$$F = x' F_{x'} + y' F_{y'}, \tag{3}$$

which makes the integral J independent of the particular parameter representation chosen for the curve (1). In order to avoid complication it will further be assumed that the problem is a *regular* one, i. e., that the function

$$F_1 = \frac{1}{y'^2} F_{x'x'} = -\frac{1}{x'y'} F_{x'y'} = \frac{1}{x'^2} F_{y'y'} \tag{4}$$

is everywhere different from zero.

The functions $a(u)$, $b(u)$ defining the curve L are of class C'.

The EULER equations which define the extremals for the integral J have the form

$$F_x - \frac{d}{dt} F_{x'} = 0, \qquad F_y - \frac{d}{dt} F_{y'} = 0, \tag{5}$$

---

* In a physical problem F or f will usually be defined only on one side of L. In order to satisfy this assumption their definitions may be extended arbitrarily over the whole of R.

† By a function of class $C^{(n)}$ is meant one which with all its derivatives of order up to and including n, is continuous.

‡ For an explanation of the notions and well known theorems of the calculus of variations used in the following pages, the reader is referred to BOLZA, *Lectures on the Calculus of Variations*, Chicago, 1904.

and these are equivalent to the single equation

$$F_{xy'} - F_{x'y} + F_1(x'y'' - x''y') = 0. \tag{6}$$

Equation (6) is equivalent to the system

$$\frac{dx}{ds} = \cos\Gamma, \qquad \frac{dy}{ds} = \sin\Gamma, \qquad \frac{d\Gamma}{ds} = \frac{F_{x'y} - F_{xy'}}{F_1},$$

where s is the length of arc and Γ the direction of the tangent on the curve. In the third equation the arguments of $F_{x'y}$, $F_{xy'}$, F_1, are x, y, $\cos\Gamma$, $\sin\Gamma$.

The assumptions made on F and the existence theorems * on differential equations, justify the following statement:

Through every point (x_0, y_0) of the region R there passes one and but one extremal for the integral J in the direction Γ_0. These curves extend from boundary to boundary of the region R. Analytically they may be represented by the equations

$$x = \Phi(s, x_0, y_0, \Gamma_0), \qquad y = \Psi(s, x_0, y_0, \Gamma_0), \tag{7}$$

where Φ and Ψ satisfy the initial conditions,

$$x_0 = \Phi(0, x_0, y_0, \Gamma_0), \qquad y_0 = \Psi(0, x_0, y_0, \Gamma_0),$$

$$\cos\gamma_0 = \Phi_s(0, x_0, y_0, \Gamma_0), \qquad \sin\gamma_0 = \Psi_s(0, x_0, y_0, \Gamma_0).$$

A similar set of extremals exists for the integral j, which will be denoted by

$$x = \phi(s, x_0, y_0, \gamma_0), \qquad y = \psi(s, x_0, y_0, \gamma_0). \tag{8}$$

The functions Φ and Ψ have all first partial derivatives, and the particular second derivatives Φ_{ss}, $\Phi_{s\Gamma_0}$, Ψ_{ss}, $\Psi_{s\Gamma_0}$, continuous. This property is true for any (x_0, y_0) in R, Γ_0 any real value, and s any value defining a point (x, y) in R. The length of the interval within which s may vary depends of course upon the choice of the initial values x_0, y_0, Γ_0.

Two of the conditions upon the curve C_{123} which minimizes the sum of the integrals J_{12} and j_{23} are easily derived from the usual theory. The arcs C_{12} and C_{23} must minimize J and j with respect to other curves joining the point 1 with the point 2 and 2 with 3 respectively, so that

I. *C_{12} must be an extremal for the integral J, i. e., one of the set (7). C_{23} must be an extremal for the integral j, i. e., one of the set (8).*

II. *If C_{12} and C_{23} satisfy condition* I, *then, according to Legendre, the functions F_1 and f_1 must be positive along C_{12} and C_{23}, in order that there may be a minimum and not a maximum.*

On account of the assumption that the problem is regular the WEIERSTRASS

*See BLISS, Annals of Mathematics, ser. 2, vol. 6 (1905), p. 49.

E function is never negative. The minimum, if one exists, will therefore be a strong minimum.

The third condition is a restriction upon the directions of the extremals C_{12} and C_{23} at their intersection with the curve L.

III. *The directions* Γ, γ, θ, *of the extremals* C_{12}, C_{23} *and the curve* L *respectively, at their intersection point* $2(x_2, y_2)$, *must satisfy the relation*

$$\begin{aligned} &F_{x'}(x_2, y_2, \cos\Gamma, \sin\Gamma)\cos\theta + F_{y'}(x_2, y_2, \cos\Gamma, \sin\Gamma)\sin\theta \\ &\quad = f_{x'}(x_2, y_2, \cos\gamma, \sin\gamma)\cos\theta + f_{y'}(x_2, y_2, \cos\gamma, \sin\gamma)\sin\theta. \end{aligned} \tag{9}$$

A similar condition has been proved by BOLZA* in the problem which corresponds to reflection in the way that the problem here considered corresponds to refraction. In the case of reflection but one integral J is involved. The proof in the present case is, however, so similar that it need not be given in detail. It consists in setting up a family of curves depending upon a parameter α, the curve C_{123} being given for the value $\alpha = 0$. Each of the curves joins 1 and 3 and crosses L once at a point $2'$. The sum of the integrals, $J_{12'} + j_{2'3}$, taken along one of these curves, is a function of α, and its derivative must vanish for $\alpha = 0$. After an integration by parts, well known in the calculus of variations, the integrals in the expression for the derivative vanish because C_{12} and C_{23} are extremals, and the condition (9) results.

In the problem of the refraction of a light ray at the surface between two homogeneous media the functions F and f are

$$\frac{\sqrt{x'^2 + y'^2}}{V}, \qquad \frac{\sqrt{x'^2 + y'^2}}{v},$$

where V and v are the constant velocities of light in the two media. Equation (9) becomes

$$\frac{\cos(\Gamma - \theta)}{V} = \frac{\cos(\gamma - \theta)}{v},$$

or

$$\frac{\sin i}{\sin r} = \frac{v}{V},$$

where i and r are the angles of incidence and refraction respectively.

§ 2. *The Jacobi condition.*

The following proof of the fourth necessary condition is geometrical and does not cover some cases in which the envelope of the set of extremals used has a singular point of a special kind. This objection holds also for the proof in

*BOLZA, loc. cit., p. 151.

geometrical form given originally by KNESER* for the ordinary problem of the calculus of variations. The method has the advantage, however, of being relatively simple, and one who is familiar with the analytical discussion of the second variation will readily find the more comprehensive analytical proof. The proof will be based upon the properties of a particular family of curves consisting of a set of extremals for J and a set for j. The existence of these sets is a consequence of the statement on page 327.

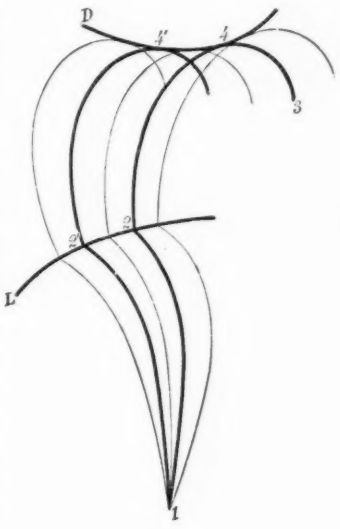


FIG. 1.

Suppose that C_{123} is a curve joining the points 1 and 3, intersecting L at the point 2, and satisfying the conditions I, II, III of the preceding section. Let $C_{12'}$ be a neighboring extremal to C_{12} through the point 1 intersecting the curve L at a point $2'$ (see figure 1). It will be shown that by equation (9) the direction of $C_{12'}$ at $2'$ defines uniquely another direction at that point, through which a new extremal for the integral j may be passed. The one parameter set of extremals for J through the point 1 defines in this way another one-parameter set for j having initial points on the curve L. It is to be proved that the arc C_{123} cannot minimize the sum of the two integrals if it contains a point of contact with an envelope of this second set.

Consider the extremals

$$(10)\quad C_{12'}: \qquad x = \Phi(s, x_1, y_1, \Gamma), \qquad y = \Psi(s, x_1, y_1, \Gamma)$$

with parameter Γ, formed from equations (7) by giving x_0, y_0 the values of the coördinates of the point 1. For the present it will be assumed that the point 2 is not conjugate to the point 1, so that the determinant

$$\Delta = \begin{vmatrix} \Phi_s & \Phi_\Gamma \\ \Psi_s & \Psi_\Gamma \end{vmatrix}$$

is different from zero for the values of s and Γ defining the point 2.

It will be convenient to use the value of u, the parameter of the curve L, at the point $2'$ as the parameter of the family instead of Γ. The change may be effected by solving the equations

$$\Phi(s, x_1, y_1, \Gamma) = a(u), \qquad \Psi(s, x_1, y_1, \Gamma) = b(u)$$

* Mathematische Annalen, vol. 50 (1898), p. 27. Compare with his *Lehrbuch der Variationsrechnung* (Braunschweig, 1900), p. 93.

for s and Γ in terms of u and substituting the value of Γ in the equations (10). The usual theorems on implicit functions show that s and Γ are defined as functions of u of class C', since by the hypothesis on the extremal C_{12} the above equations have a set of solutions corresponding to the point 2, and the determinant Δ is different from zero at 2. After substituting the value of Γ, the equations (10) take the form*

$$(11) \qquad C_{12'}: \qquad x = \Phi(s, u), \qquad y = \Psi(s, u).$$

In the equation (9) x_2, y_2, θ may be expressed by the functions of u which define the curve L, and Γ by the function of u derived from (11) which defines the directions of the extremals $C_{12'}$ at the points $2'$. The equation (9) formed in this way for the point $2'$ is then one between u and γ which by hypothesis has one solution corresponding to the point 2. Furthermore the derivative for γ is easily found with the help of equations (4) to be

$$f_1 \sin(\theta - \gamma),$$

which is different from zero at the point 2, unless C_{23} is tangent to L. Therefore γ is defined as a function of u of class C', unless C_{23} is tangent to L, and the set of extremals for j with initial points on L is given by the equations

$$(12) \qquad x = \phi(s, u), \qquad y = \psi(s, u),$$

formed by substituting $a(u)$, $b(u)$, $\gamma(u)$ for x_0, y_0, γ in equations (8).

In case C_{23} is tangent to L we could reverse the procedure, starting from a set of extremals for j through 2, and determining a set for J with initial points on L, by means of equation (9). Therefore, unless both C_{12} and C_{23} are tangent to L at 2, it may be assumed without restriction that C_{23} is the one not tangent, and the above development holds. *It will be assumed that this singular case does not occur for the problem under consideration, i. e., that C_{12} and C_{23} are not both tangent to L.*

The results obtained may be restated in the following theorem:

If C_{123} is a curve satisfying the conditions I, II, III, *then to each extremal $C_{12'}$ of the integral J near C_{12} there corresponds an extremal of the integral j which, with $C_{12'}$, satisfies the corner condition* III *at the point* $2'$. *The one-parameter set of extremals of J through the point* 1, *defines in this way another one-parameter set belonging to j with initial points on the curve L. The equations of the two sets may be put in the forms* (11) *and* (12) *with u as parameter.*

Suppose that the extremals (12) have an envelope D which touches the extremal $C_{2'4'}$ at the point $4'$, and the extremal C_{23} at 4, and suppose that the direction $2'4'$ on $C_{2'4'}$ coincides with $4'4$ on D† (see figure 1).

* For convenience the same symbols Φ and Ψ are used for the equations in this new form.

† The envelope D will always have a branch $4'4$ coinciding in direction with $2'4'$ unless it has a cusp of special kind at the point 4. See p. 332.

The integral J taken along $C_{12'}$, plus the integral j taken along $C_{2'4'}$ and $D_{4'1}$, apparently depends upon the choice of the extremal $C_{12'}$, that is, upon the parameter u. But in reality this sum is a constant, as may be shown by forming its derivative with respect to u.

The derivative of $J_{12'}$ is

$$\frac{dJ_{12'}}{du} = \frac{d}{du}\int_1^{2'} F(\Phi, \Psi, \Phi', \Psi')\,ds$$

$$= F\frac{ds}{du}\Big|^{2'} + \int_1^{2'} \{F_x\Phi_u + F_y\Psi_u + F_{x'}\Phi'_u + F_{y'}\Psi'_u\}\,ds$$

$$= F\frac{ds}{du} + F_{x'}\Phi_u + F_{y'}\Psi_u\Big|^{2'} + \int_1^{2'} \{(F_x - F'_{x'})\Phi_u + (F_y - F'_{y'})\Psi_u\}ds,$$

since Φ_u and Ψ_u vanish for $s = 0$. This follows because all the extremals (11) pass through the point 1 for $s = 0$, i. e.,

$$x_1 = \Phi(0, u), \qquad y_1 = \Psi(0, u).$$

The integral in the above equation vanishes on account of (5) which is satisfied by the extremals (11). The remainder of the right member may be transformed by equation (3), so that

$$\frac{dJ_{12'}}{du} = \left(\Phi'\frac{ds}{du} + \Phi_u\right)F_{x'} + \left(\Psi'\frac{ds}{du} + \Psi_u\right)F_{y'}\Big|^{2'}.$$

Now the value of s corresponding to the point $2'$ is the solution of the equations

$$\Phi(s, u) = a(u), \qquad \Psi(s, u) = b(u),$$

and hence

$$\frac{dJ_{12'}}{du} = F_{x'}(\Phi, \Psi, \Phi', \Psi')\frac{da}{du} + F_{y'}(\Phi, \Psi, \Phi', \Psi')\frac{db}{du}\Big|^{2'}. \tag{13}$$

In a similar manner it is found that

$$\frac{dj_{2'4'}}{du} = \frac{d}{du}\int_{2'}^{4'} f(\phi, \psi, \phi', \psi')\,ds$$

$$= -f_{x'}\frac{da}{du} - f_{y'}\frac{db}{du}\Big|^{2'} + f_{x'}\frac{d\lambda}{du} + f_{y'}\frac{d\mu}{du}^{4'}, \tag{14}$$

where $\lambda(u)$, $\mu(u)$ are the functions defining x and y for the envelope D. In deriving this result it should be remembered that the extremals (12) intersect L for the parameter value $s = 0$, a fact expressed by the equations

$$\phi(0, u) = a(u), \qquad \psi(0, u) = b(u);$$

and furthermore that λ and μ are defined by the equations

$$\lambda(u) = \phi(s, u), \qquad \mu(u) = \psi(s, u),$$

in which s is the solution of

$$\delta = \begin{vmatrix} \phi_s & \phi_u \\ \psi_s & \psi_u \end{vmatrix} = 0$$

in terms of u.

If the envelope D has not a singular point at 4, the expression (14) may be put into a more convenient form. Since

$$\frac{d\lambda}{du} = \phi_s \frac{ds}{du} + \phi_u, \qquad \frac{d\mu}{du} = \psi_s \frac{ds}{du} + \Psi_u,$$

where s is the solution of $\delta = 0$, it follows readily that

$$\begin{vmatrix} \phi_s & \dfrac{d\lambda}{du} \\ \psi_s & \dfrac{d\mu}{du} \end{vmatrix} = 0.$$

Since the extremal is assumed to be free of singular points ϕ_s and ψ_s cannot vanish together. Unless D has a singular point at 4 the functions $d\lambda/du$, $d\mu/du$ do not vanish together in the neighborhood of the point 4, and there exists a function κ of constant sign such that

$$\phi_u = \kappa \frac{d\lambda}{du}, \qquad \psi_u = \kappa \frac{d\mu}{du}.$$

It may be assumed that $\kappa > 0$, for changing u to $-u$ changes the sign of κ. Substituting these values of ϕ_u and ψ_u in the last two terms of (14) and applying (2) and (3) there results:

$$\frac{dj_{x'x'}}{du} = -f_{x'}(\phi, \psi, \phi', \psi')\frac{du}{du} - f_{y'}(\phi, \psi, \phi', \psi')\frac{db}{du}^{x'} + f(\lambda, \mu, \lambda', \mu')^{x'}. \tag{15}$$

In case the envelope D has a singular point at 4′ the above transformation of equation (14) may not be possible, and the proof given here may fail.

Finally

$$\frac{dj_{x'x}}{du} = \frac{d}{du}\int_{x'}^{x} f(\lambda, \mu, \lambda', \mu')du$$

$$= -f(\lambda, \mu, \lambda', \mu')|^{x'}. \tag{16}$$

The sum of the three expressions (13), (15) and (16) vanishes identically on

account of equation (9). The theorem corresponding to KNESER's theorem in the ordinary theory follows at once:

If the envelope D of the extremals $C_{2'4'}$ has not a singular point at its point of contact with C_{23}, then the value of the sum of the integrals $J_{12'}$ taken along the extremal $C_{12'}$ from the point 1 to the curve L, $j_{2'4'}$ taken along the extremal $C_{2'4'}$ from the curve L to the point of contact of $C_{2'4'}$ with the envelope D, and $j_{4'4}$ taken along D from the point of contact with $C_{2'4'}$ to the point of contact with C_{24}, always has the same value as the sum of J_{12} and j_{24} taken along the extremals C_{12} and C_{24} respectively (see figure 1).

With this theorem in mind it is easy to see that the extremal C_{123} cannot minimize $J_{12}+j_{23}$ if the point 4 lies on the arc C_{23} and is not a singular point of the envelope D. For D cannot be an extremal, since under the assumptions there exists but one extremal through a given point in a given direction, and therefore the arc of D between $4'$ and 4 could be replaced by another curve giving a smaller value to j. It follows that 1 and 3 could be joined by a curve through a point $2'$ of L, giving $J_{12'}+j_{2'3}$ a smaller value than

$$J(C_{12'})+j(C_{2'4'})+j(D_{4'4})+j(C_{2'3}),$$

and hence smaller than $J(C_{12})+j(C_{23})$.

The particular cases where 1 and 2, or 2 and 3 are conjugate points on C_{12} or C_{23} need special discussion. If 2 and 3 are conjugate the point 4 lies between them. For the determinant δ is a solution of JACOBI's differential equation, a linear homogeneous differential equation of second order, having in this case, since 2 and 3 are conjugate, a solution which vanishes at 2 and 3. By the well known STURM's theorems concerning the separation of the roots of such solutions, δ will have a zero between 2 and 3, and C_{123} will not minimize the integral. If 1 and 2 were conjugate in C_{12}, but 2 and 3 not conjugate in C_{23}, the start should be made with the one-parameter set of extremals through 3 instead of through 1, and the same conclusion would be reached — C_{123} is not a minimizing curve. In case both 1, 2 and 2, 3 were conjugate pairs, a point $1'$ on C_{12} between 1 and 2 could be used as a starting point instead of 1, and it would follow that the contact point 4 would again be between 2 and 3, since they are conjugate, and C_{123} would not minimize $J+j$. To summarize:

IV. *The extremal arcs C_{12} and C_{23}, including the end points, can contain no points conjugate to their initial points* 1 *and* 2, *and the envelope of the extremals $C_{2'4'}$ must not touch the arc C_{23} between* 2 *and* 3.*

*The proof given above shows that the envelope cannot touch C_{23}, even at the point 3, unless there is a singular point at 3. On the other hand the proof fails in case the envelope touches between 2 and 3 and has a singular point. The condition in this case may, however, be derived analytically by consideration of the second variation.

§ 3. *The sufficient conditions.*

Suppose a curve C_{123} exists which crosses L but once (at 2) without having both arcs C_{12} and C_{23} tangent to L, and satisfies the conditions I, II, III, of the preceding sections besides the condition

IV′. *The extremal arcs C_{12} and C_{23}, including the end points, contain no points conjugate to their initial points* 1 *and* 2, *and the envelope of the extremals $C_{2'4'}$ does not touch the arc C_{23} between* 2 *and* 3 OR AT 3.

It will be proved by means of HILBERT's invariant integral that the curve C_{123} minimizes the sum of the integrals $J + j$.

As has been shown above, the set of extremals $C_{12'}$ through 1 determines a set $C_{2'4'}$ for the integral j which, with $C_{12'}$, satisfy the corner condition III. Since by condition IV′ the determinants Δ and δ for these two sets are different from zero between 1 and 2 and at 2, and between 2 and 3 and at 3 respectively, it follows that the extremals

$$x = \Phi(s, u), \qquad y = \Psi(s, u), \tag{17}$$

for the integral J, and the extremals

$$x = \phi(s, u), \qquad y = \psi(s, u), \tag{18}$$

for the integral j, form two fields, B and b, about the arcs C_{12} and C_{23} respectively. In B and b the derivatives Φ', Ψ'; ϕ', ψ', which determine the directions of the extremals, are functions of x and y.

Now the integral

$$J^* = \int \{ F_{x'}(x, y, \Phi', \Psi')x' + F_{y'}(x, y, \Phi', \Psi')y' \}\, dt,$$

where Φ', Ψ' are these functions of x and y, taken along any path

$$x = x(t), \qquad y = y(t),$$

which lies in the field B, is independent of the path and depends only on the end points.* Furthermore it follows from the homogeneity of F that J^* taken along an extremal is equal to J. Similarly the integral

$$j^* = \int \{ f_{x'}(x, y, \phi', \psi')x' + f_{y'}(x, y, \phi', \psi')y' \}\, dt$$

is independent of the path in b and reduces to j when the path is an extremal for the integral j.

Consider any curve C,

$$C: \qquad x = x(t), \qquad y = y(t),$$

* BOLZA, loc. cit., p. 195.

which crosses L once at the point **6**, lies entirely in the two fields B and b, and joins two points in B and b respectively. *The sum* $J^*(C_{56})+j^*(C_{67})$ *is independent of the path and depends only on the end points* 5 *and* 7. In fact, let $\bar{C}$ be any other curve joining 5 and 7, and crossing L at 6′ (see figure 2):

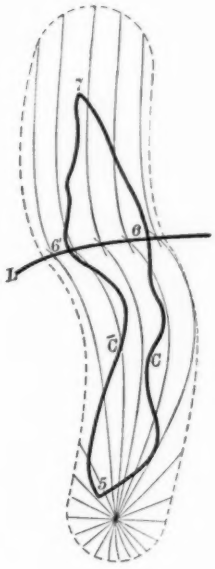


FIG. 2.

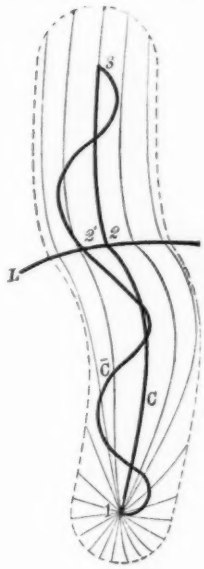


FIG. 3.

then, since the integrals J^* and j^* are invariant in the fields B and b respectively,

$$J^*(\bar{C}_{56'})+J^*(L_{6'6})=J^*(C_{56}),$$

$$j^*(L_{66'})+j^*(\bar{C}_{6'7})=j^*(C_{67}).$$

Therefore,

$$J^*(\bar{C}_{56'})+j^*(\bar{C}_{6'7})-[J^*(C_{56})+j^*(C_{67})]=-J^*(L_{6'6})-j^*(L_{66'})$$
$$=J^*(L_{66'})-j^*(L_{66'}),$$

since the integrand of J^* changes sign when the direction of integration is reversed. Hence

$$J^*(\bar{C}_{56'})+j^*(\bar{C}_{6'7})-[J^*(C_{56})+j^*(C_{67})]$$
$$=\int_6^{6'}\Big\{F_{x'}(a,b,\Phi',\Psi')\frac{da}{du}+F_{y'}(a,b,\Phi',\Psi')\frac{db}{du}$$
$$-f_{x'}(a,b,\phi',\psi')\frac{da}{du}-f_{y'}(a,b,\phi',\psi')\frac{db}{du}\Big\}du.$$

and furthermore that λ and μ are defined by the equations

$$\lambda(u) = \phi(s, u), \qquad \mu(u) = \psi(s, u),$$

in which s is the solution of

$$\delta = \begin{vmatrix} \phi_s & \phi_u \\ \psi_s & \psi_u \end{vmatrix} = 0$$

in terms of u.

If the envelope D has not a singular point at 4, the expression (14) may be put into a more convenient form. Since

$$\frac{d\lambda}{du} = \phi_s \frac{ds}{du} + \phi_u, \qquad \frac{d\mu}{du} = \psi_s \frac{ds}{du} + \Psi_u,$$

where s is the solution of $\delta = 0$, it follows readily that

$$\begin{vmatrix} \phi_u & \dfrac{d\lambda}{du} \\ \psi_s & \dfrac{d\mu}{du} \end{vmatrix} = 0.$$

Since the extremal is assumed to be free of singular points ϕ_s and ψ_s cannot vanish together. Unless D has a singular point at 4 the functions $d\lambda/du$, $d\mu/du$ do not vanish together in the neighborhood of the point 4, and there exists a function κ of constant sign such that

$$\phi_s = \kappa \frac{d\lambda}{du}, \qquad \psi_s = \kappa \frac{d\mu}{du}.$$

It may be assumed that $\kappa > 0$, for changing u to $-u$ changes the sign of κ. Substituting these values of ϕ_s and ψ_s in the last two terms of (14) and applying (2) and (3) there results:

$$\frac{dj_{2'4'}}{du} = -f_{x'}(\phi, \psi, \phi', \psi') \frac{da}{du} - f_{y'}(\phi, \psi, \phi', \psi') \frac{db}{du}^{2'} + f(\lambda, \mu, \lambda', \mu')^{4'}. \tag{15}$$

In case the envelope D has a singular point at 4' the above transformation of equation (14) may not be possible, and the proof given here may fail.

Finally

$$\frac{dj_{4'1}}{du} = \frac{d}{du}\int_{4'}^{1} f(\lambda, \mu, \lambda', \mu')\, du$$

$$= -f(\lambda, \mu, \lambda', \mu')|^{4'}. \tag{16}$$

The sum of the three expressions (13), (15) and (16) vanishes identically on

account of equation (9). The theorem corresponding to KNESER's theorem in the ordinary theory follows at once:

If the envelope D of the extremals $C_{2'4'}$ has not a singular point at its point of contact with C_{23}, then the value of the sum of the integrals $J_{12'}$ taken along the extremal $C_{12'}$ from the point 1 to the curve L, $j_{2'4'}$ taken along the extremal $C_{2'4'}$ from the curve L to the point of contact of $C_{2'4'}$ with the envelope D, and $j_{4'4}$ taken along D from the point of contact with $C_{2'4'}$ to the point of contact with C_{24}, always has the same value as the sum of J_{12} and j_{24} taken along the extremals C_{12} and C_{24} respectively (see figure 1).

With this theorem in mind it is easy to see that the extremal C_{123} cannot minimize $J_{12} + j_{23}$ if the point 4 lies on the arc C_{23} and is not a singular point of the envelope D. For D cannot be an extremal, since under the assumptions there exists but one extremal through a given point in a given direction, and therefore the arc of D between 4′ and 4 could be replaced by another curve giving a smaller value to j. It follows that 1 and 3 could be joined by a curve through a point 2′ of L, giving $J_{12'} + j_{2'3}$ a smaller value than

$$J(C_{12'}) + j(C_{2'4'}) + j(D_{4'4}) + j(C_{24}),$$

and hence smaller than $J(C_{12}) + j(C_{23})$.

The particular cases where 1 and 2, or 2 and 3 are conjugate points on C_{12} or C_{23} need special discussion. If 2 and 3 are conjugate the point 4 lies between them. For the determinant δ is a solution of JACOBI's differential equation, a linear homogeneous differential equation of second order, having in this case, since 2 and 3 are conjugate, a solution which vanishes at 2 and 3. By the well known STURM's theorems concerning the separation of the roots of such solutions, δ will have a zero between 2 and 3, and C_{123} will not minimize the integral. If 1 and 2 were conjugate in C_{12}, but 2 and 3 not conjugate in C_{23}, the start should be made with the one-parameter set of extremals through 3 instead of through 1, and the same conclusion would be reached — C_{123} is not a minimizing curve. In case both 1, 2 and 2, 3 were conjugate pairs, a point 1′ on C_{12} between 1 and 2 could be used as a starting point instead of 1, and it would follow that the contact point 4 would again be between 2 and 3, since they are conjugate, and C_{123} would not minimize $J + j$. To summarize:

IV. *The extremal arcs C_{12} and C_{23}, including the end points, can contain no points conjugate to their initial points* 1 *and* 2, *and the envelope of the extremals $C_{2'4'}$ must not touch the arc C_{23} between* 2 *and* 3.*

* The proof given above shows that the envelope cannot touch C_{23}, even at the point 3, unless there is a singular point at 3. On the other hand the proof fails in case the envelope touches between 2 and 3 and has a singular point. The condition in this case may, however, be derived analytically by consideration of the second variation.

§ 3. *The sufficient conditions.*

Suppose a curve C_{123} exists which crosses L but once (at 2) without having both arcs C_{12} and C_{23} tangent to L, and satisfies the conditions I, II, III, of the preceding sections besides the condition

IV′. *The extremal arcs C_{12} and C_{23}, including the end points, contain no points conjugate to their initial points* 1 *and* 2, *and the envelope of the extremals $C_{2'4'}$ does not touch the arc C_{23} between* 2 *and* 3 OR AT 3.

It will be proved by means of HILBERT's invariant integral that the curve C_{123} minimizes the sum of the integrals $J+j$.

As has been shown above, the set of extremals $C_{12'}$ through 1 determines a set $C_{2'4'}$ for the integral j which, with $C_{12'}$, satisfy the corner condition III. Since by condition IV′ the determinants Δ and δ for these two sets are different from zero between 1 and 2 and at 2, and between 2 and 3 and at 3 respectively, it follows that the extremals

$$x=\Phi(s,u), \qquad y=\Psi(s,u), \tag{17}$$

for the integral J, and the extremals

$$x=\phi(s,u), \qquad y=\psi(s,u), \tag{18}$$

for the integral j, form two fields, B and b, about the arcs C_{12} and C_{23} respectively. In B and b the derivatives Φ', Ψ'; ϕ', ψ', which determine the directions of the extremals, are functions of x and y.

Now the integral

$$J^*=\int\{F_{x'}(x,y,\Phi',\Psi')x'+F_{y'}(x,y,\Phi',\Psi')y'\}\,dt,$$

where Φ', Ψ' are these functions of x and y, taken along any path

$$x=x(t), \qquad y=y(t),$$

which lies in the field B, is independent of the path and depends only on the end points.* Furthermore it follows from the homogeneity of F that J^* taken along an extremal is equal to J. Similarly the integral

$$j^*=\int\{f_{x'}(x,y,\phi',\psi')x'+f_{y'}(x,y,\phi',\psi')y'\}\,dt$$

is independent of the path in b and reduces to j when the path is an extremal for the integral j.

Consider any curve C,

$$C: \qquad x=x(t), \qquad y=y(t),$$

* BOLZA, loc. cit., p. 195.

which crosses L once at the point **6**, lies entirely in the two fields B and b, and joins two points in B and b respectively. *The sum* $J^*(C_{56})+j^*(C_{67})$ *is independent of the path and depends only on the end points* 5 *and* 7. In fact, let $\bar{C}$ be any other curve joining 5 and 7, and crossing L at 6′ (see figure 2):

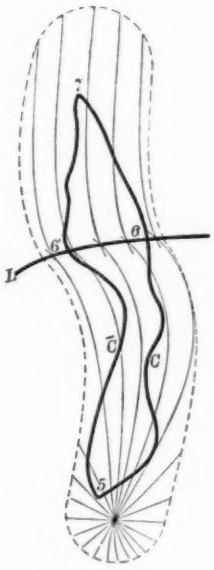


FIG. 2.

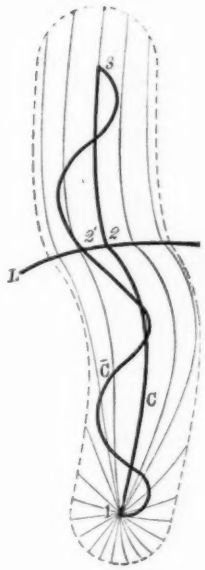


FIG. 3.

then, since the integrals J^* and j^* are invariant in the fields B and b respectively,

$$J^*(\bar{C}_{56'})+J^*(L_{6'6})=J^*(C_{56}),$$
$$j^*(L_{66'})+j^*(\bar{C}_{6'7})=j^*(C_{67}).$$

Therefore,

$$J^*(\bar{C}_{56'})+j^*(\bar{C}_{6'7})-[J^*(C_{56})+j^*(C_{67})]=-J^*(L_{6'6})-j^*(L_{66'})$$
$$=J^*(L_{66'})-j^*(L_{66'}),$$

since the integrand of J^* changes sign when the direction of integration is reversed. Hence

$$J^*(\bar{C}_{56'})+j^*(\bar{C}_{6'7})-[J^*(C_{56})+j^*(C_{67})]$$
$$=\int_6^{6'}\left\{F_{x'}(a,b,\Phi',\Psi')\frac{da}{du}+F_{y'}(a,b,\Phi',\Psi')\frac{db}{du}\right.$$
$$\left.-f_{x'}(a,b,\phi',\psi')\frac{da}{du}-f_{y'}(a,b,\phi',\psi')\frac{db}{du}\right\}du.$$

Since the extremals (17), (18) satisfy equation (9), the integrand of the right member is identically zero and $J^* + j^*$ is independent of the path. Furthermore, if the curve C_{567} coincides with an extremal for J from 5 to 6, and with an extremal for j from 6 to 7, then

$$J^*(C_{56}) + j^*(C_{67}) = J(C_{56}) + j(C_{67}).$$

Let $\bar{C}$ (figure 3) be any curve joining the points 1 and 2, lying entirely in the two fields B and b, and crossing L at $2'$, and let C be the curve satisfying conditions I, II, III, IV′. Then

$$\begin{aligned} J^*(\bar{C}_{12'}) + j^*(\bar{C}_{2'3}) &= J^*(C_{12}) + j^*(C_{23}) \\ &= J(C_{12}) + j(C_{23}). \end{aligned}$$

Therefore, by substituting this value of $J(C_{12}) + j(C_{23})$,

$$\begin{aligned} J(\bar{C}_{12'}) + j(\bar{C}_{2'3}) - [J(C_{12}) + j(C_{23})] &= J(\bar{C}_{12'}) - J^*(\bar{C}_{12'}) + j(\bar{C}_{2'3}) - j^*(\bar{C}_{2'3}) \\ &= \int_{\bar{C}_{12'}} E dt + \int_{\bar{C}_{23'}} e dt, \end{aligned}$$

where E and e are the WEIERSTRASS E functions for the fields B and b respectively. But as was stated in § 1 the E functions are nowhere negative, and they vanish only when the direction of $\bar{C}$ at any point coincides with the direction of the extremal at that point. Therefore

$$J(\bar{C}_{12'}) + j(\bar{C}_{2'3}) > J(C_{12}) + j(C_{23}),$$

and the curve C_{123} minimizes the sum of the integrals. The result may be stated in the following theorem:

A curve C_{123} which joins the fixed points 1, 3, *and crosses the curve L once at the point* 2 *without having both branches tangent to L at* 2, *minimizes the sum of the integrals J and j if it satisfies the conditions I, II, III, IV′.*

The above discussion may be readily extended to cover the case of any number of curves of discontinuity for the integrand of the calculus of variations problem, i. e., in the case of light, to any number of refracting surfaces. The methods used above apply also with but slight modifications to the problem of the calculus of variations corresponding to the reflection of light.